

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 6478 Seconds  
(without alignments)  
11278.571 Million cell updates/sec

Title: US-09-508-658A-3

Perfect score: 1545  
Sequence: 1 agagaagctgaggtctctc.....aaaataataaattagctg 1545

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 200 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_rst:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541.8	99.8	1545	9	AB006683 Homo sapi
2	1540.2	99.7	1545	6	CQ731863 Sequence
3	1367.8	88.5	1463	9	AB006685 Homo sapi
4	1023.6	66.3	2027	9	AB006682 Homo sapi
5	1023.6	66.3	2245	6	A97284 Sequence 1
6	1023.6	66.3	2245	6	HSAPCEDD
7	1023.6	66.3	2258	6	CQ731864 Sequence
8	462.4	29.9	20000	9	AB006684 Homo sapi
9	462.4	29.9	36284	9	HSAP9610 Homo sapi
10	462.4	29.9	42133	9	AP001060 Homo sapi
11	462.4	29.9	110000	2	AC003656_5
12	462.4	29.9	340000	9	AP001754 Homo sapi
13	451.2	29.2	143192	9	AL954247 Pan trogl
14	385.6	25.0	1656	10	AF128772 Mus muscu
15	385.6	25.0	1656	10	AF128773 Mus muscu
16	385.6	25.0	1659	6	A97293 Sequence 10
17	385.6	25.0	1659	10	MMU132243 Mus muscu
18	385.6	25.0	1906	10	MMU24821 Mus muscu
19	385.6	25.0	1924	10	AF128116 Mus muscu

20	385.6	25.0	1936	10	AF079536 Mus muscu
21	375.6	24.3	1921	10	AF128117 Mus muscu
22	375.6	24.3	1933	10	AF128115 Mus muscu
23	336	21.7	2586	9	HSJTPK
24	174.6	11.3	1744	10	AF128121 Mus muscu
25	174.6	11.3	1747	10	AF128120 Mus muscu
26	174.6	11.3	1756	10	AF128119 Mus muscu
27	174.6	11.3	1759	10	AF128118 Mus muscu
28	159	10.3	1625	10	AF128124 Mus muscu
29	159	10.3	1637	10	AF128122 Mus muscu
30	149	9.6	1622	10	AF128125 Mus muscu
31	149	9.6	1634	10	AF128123 Mus muscu
32	118.4	7.7	9876	6	AX344994 Sequence
33	101.8	6.6	9876	6	AX344995 Sequence
34	96.6	6.3	581	11	BV089377 RPAMSEQ0
35	96.6	6.3	18351	10	AF105002 Mus muscu
36	96.6	6.3	18616	10	MMU007715 Mus muscu
37	96.6	6.3	46872	10	AF073797 Mus muscu
38	96.6	6.3	158049	2	AC015891 Mus muscu
39	96.6	6.3	190019	2	AC138672 Mus muscu
40	90.2	5.8	217688	2	AC108592 Ratius no
41	90.2	5.8	283951	2	AC109383 Ratius no
42	82.6	5.3	558	11	BV163805 RPAMSEQ0
43	82.6	5.3	574	11	BV089350 RPAMSEQ0
44	82.6	5.3	593	11	BV088992 RPAMSEQ0
45	77.2	5.0	125020	9	AF429315 Homo sapi
46	73.6	4.8	125020	9	AF429315 Homo sapi
47	70	4.5	184594	5	BK649390 Zebrafish
48	65	4.2	9466	9	AK127046 Homo sapi
49	65	4.2	9466	9	AF425231 Homo sapi
50	63.2	4.1	6438	10	BC058578 Mus muscu
51	63.2	4.1	67573	2	AC122760 Mus muscu
52	63.2	4.1	173909	2	AC134529 Mus muscu
53	62.8	4.0	572	11	BV089378 RPAMSEQ0
54	61.4	4.0	248835	2	AC097256 Ratius no
55	61.4	4.0	273874	2	AC127615 Ratius no
56	60.4	3.9	121375	5	BX004850 Zebrafish
57	60	3.9	60	6	CQ641728 Homo sapi
58	60	3.9	6327	6	A56817 Sequence 1
59	60	3.9	6328	6	AR224019 Sequence
60	60	3.9	6328	6	AR269336 Sequence
61	60	3.9	6328	6	AR367436 Sequence
62	60	3.9	6412	6	CQ722409 Sequence
63	60	3.9	6417	6	AX332238 Sequence
64	60	3.9	6417	6	HSMT1218 Sequence
65	60	3.9	6475	6	AR338834 Sequence
66	60	3.9	6584	6	BC038596 Homo sapi
67	60	3.9	132592	2	AL589792 Homo sapi
68	60	3.9	166518	9	HS120622 Human DNA
69	59.8	3.9	1053	6	AR383186 Sequence
70	59.8	3.9	1591	9	HSU08379 Human M1-2
71	59.8	3.9	6331	6	CQ491372 Sequence
72	59.8	3.9	6331	6	AX924075 Sequence
73	59.8	3.9	6331	6	AX924076 Sequence
74	59.8	3.9	6331	9	AP006515 Homo sapi
75	59.8	3.9	6771	6	AX224049 Sequence
76	59.8	3.9	168479	2	AC135674 Ratius no
77	59.8	3.9	221787	2	AC1319782 Ratius no
78	59.8	3.9	221787	2	AC097926 Ratius no
79	59.8	3.9	246566	6	CQ720332 Sequence
80	59.8	3.8	2000	6	AX655393 Sequence
81	58.8	3.8	6434	9	AK128741 Homo sapi
82	58.8	3.8	6434	9	AL611985 Mouse DNA
83	58.8	3.8	219205	10	D86113 Homo sapien
84	58.2	3.8	4343	10	BC060721 Mus muscu
85	58.2	3.8	116984	2	AC133595 Homo sapi
86	58.2	3.8	187718	9	AC104581 Homo sapi
87	58.2	3.8	232984	2	AC111597 Ratius no
88	57.6	3.7	261384	2	AC115415 Ratius no
89	57.6	3.7	278068	2	AC123167 Ratius no
90	56.8	3.7	189081	5	BX511250 Zebrafish
91	56.8	3.7	2803	6	CQ721455 Sequence
92	56.6	3.7	2803	6	CQ721455 Sequence

93	56.6	3.7	6505	9	AB002296	Homo sapi
94	56.4	3.7	5811	3	AF210315	Drosophila
95	56.4	3.7	6124	6	CQ583074	Sequence
96	56.4	3.7	21747	6	CQ583073	Sequence
97	56.4	3.7	97538	2	AC017943	Drosophila
98	56.4	3.7	170939	3	AC007757	Drosophila
99	56.4	3.7	197071	3	AC007771	Drosophila
100	56.4	3.7	230001	3	AB003731	Drosophila
101	56.2	3.6	66986	2	AC133881	Homo sapi
102	56.2	3.6	165212	2	AL596125	Mouse DNA
103	56.2	3.6	183887	2	AC084415	Mus muscu
104	55.8	3.6	33957	5	AY598453	Danio rer
105	55.8	3.6	5003	6	CQ730634	Sequence
106	55.4	3.6	56550	9	AC135892	Homo sapi
107	55.4	3.6	172571	9	AC006064	Homo sapi
108	55.4	3.6	245880	2	AC079387	Homo sapi
109	54.4	3.5	135259	9	HS127820	Human DNA
110	54.2	3.5	1158	9	HST000230	Homo sapi
111	54.2	3.5	2596	9	HST000226	Homo sapi
112	54.2	3.5	3039	9	AF009353	Homo sapi
113	54.2	3.5	3799	6	CQ714282	Sequence
114	54.2	3.5	3807	9	AF119042	Homo sapi
115	54.2	3.5	3909	9	BC028689	Homo sapi
116	54.2	3.5	3997	6	CQ783904	Sequence
117	54.2	3.5	3997	6	BD127893	Primer fo
118	54.2	3.5	3997	6	AK075306	Homo sapi
119	54.2	3.5	136753	2	AC149968	Strongylo
120	53.8	3.5	300695	2	AC079431	Mus muscu
121	53.6	3.5	220674	2	BX548000	Danio rer
122	53.6	3.4	3116	5	AY598454	Danio rer
123	52.8	3.4	292	6	CQ079223	Sequence
124	52.8	3.4	292	6	CQ110776	Sequence
125	52.8	3.4	292	6	CQ149515	Sequence
126	52.8	3.4	292	6	CQ149336	Sequence
127	52.8	3.4	292	6	CQ232797	Sequence
128	52.8	3.4	292	6	CQ270820	Sequence
129	52.8	3.4	292	6	CQ308111	Sequence
130	52.8	3.4	292	6	CQ344926	Sequence
131	52.8	3.4	476	6	CQ070006	Sequence
132	52.8	3.4	476	6	CQ097609	Sequence
133	52.8	3.4	476	6	CQ136450	Sequence
134	52.8	3.4	476	6	CQ174430	Sequence
135	52.8	3.4	476	6	CQ219756	Sequence
136	52.8	3.4	476	6	CQ258219	Sequence
137	52.8	3.4	476	6	CQ295563	Sequence
138	52.8	3.4	476	6	CQ332152	Sequence
139	52.8	3.4	6025	6	CQ613956	Sequence
140	52.8	3.4	7301	2	AC018273	Drosophila
141	52.8	3.4	7303	6	CQ613955	Sequence
142	52.8	3.4	150884	2	AC009382	Drosophila
143	52.8	3.4	167816	3	AC009381	Drosophila
144	52.8	3.4	250662	3	AB003515	Drosophila
145	52.2	3.4	539	6	CQ497245	Sequence
146	52.2	3.4	539	6	CQ512372	Sequence
147	52.2	3.4	560	11	BV061719	Sequence
148	52.2	3.4	67573	2	AC122760	Sequence
149	52	3.4	1215	6	CQ715851	Sequence
150	52	3.4	1647	9	CR456448	Homo sapi
151	52	3.4	2651	6	AX833782	Sequence
152	52	3.4	2651	9	AK095777	Homo sapi
153	52	3.4	3498	6	AX747427	Sequence
154	52	3.4	3498	9	AK092243	Homo sapi
155	52	3.4	3584	6	AX747013	Sequence
156	52	3.4	3584	9	AK091480	Homo sapi
157	52	3.4	3694	9	BC012187	Homo sapi
158	51.8	3.4	3694	10	RNND4P	Homo sapi
159	51.6	3.3	4871	10	AX129293	Mus muscu
160	51.6	3.3	5686	10	AY458590	Mus muscu
161	51.4	3.3	1637	10	BC052348	Mus muscu
162	51.4	3.3	2273	10	MM748238	Mus muscu
163	51.4	3.3	212974	2	BX927371	Danio rer
164	51.4	3.3	6475	3	AF119716	Drosophila
165	51	3.3	129506	2	AC136100	Rattus no

166	50.8	3.3	155337	2	AC116408	Mus muscu
167	50.6	3.3	51	11	BV190786	Sequence
168	50.6	3.3	200	6	AX655393	Sequence
169	50.6	3.3	201768	5	BX323591	Zebrafish
170	50.4	3.3	155337	2	AC116408	Mus muscu
171	50.4	3.3	163121	5	AL928637	Zebrafish
172	50.4	3.3	170189	2	BX957358	Danio rer
173	50.2	3.2	2151	10	BC025482	Mus muscu
174	50.2	3.2	3351	10	S78219	nuclear pro
175	50.2	3.2	4032	10	BC056959	Mus muscu
176	50.2	3.2	4053	10	S78221	nuclear pro
177	50	3.2	239130	2	AC079420	Mus muscu
178	49.6	3.2	65351	2	AC139773	Homo sapi
179	49.6	3.2	170228	8	AP005829	Oryza sat
180	49.6	3.2	230387	2	AC051622	Mus muscu
181	49.4	3.2	201	11	BV179021	sgm10356
182	49.4	3.2	60032	2	AC129549	Homo sapi
183	49.4	3.2	65351	2	AC139773	Homo sapi
184	49	3.2	166862	10	AC124457	Mus muscu
185	49	3.2	167396	2	CR352259	Danio rer
186	49	3.2	202748	10	MM70307670	Mus muscu
187	49	3.2	300800	1	SC0939112	Sequence
188	48.8	3.2	3079	6	AX880162	Sequence
189	48.8	3.2	3079	6	BD158230	Primer fo
190	48.8	3.2	3079	6	AK023116	Homo sapi
191	48.8	3.2	129996	2	AC091664	Oryza sat
192	48.8	3.2	130528	8	AP003202	Oryza sat
193	48.8	3.2	136357	8	AP003562	Oryza sat
194	48.8	3.2	178331	2	BX957269	Danio rer
195	48.6	3.1	1755	6	CQ731955	Sequence
196	48.6	3.1	2251	9	BC078181	Homo sapi
197	48.6	3.1	2579	9	AY278022	Homo sapi
198	48.6	3.1	2510	9	BC039002	Homo sapi
199	48.6	3.1	2630	9	BC038222	Homo sapi
200	48.6	3.1	128266	9	AC008623	Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AB006683 1545 bp mRNA linear PRI 13-DEC-1997  
DEFINITION Homo sapiens APECD mRNA for AIRE-2, complete cde.  
ACCESSION AB006683  
VERSION AB006683.1 GI:2696616  
KEYWORDS AIRE-2; APECD; autoimmune regulator-2.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S.,  
Heino,M., Krohn,K.J.E., Lalliot,M.D., Mullis,P.E.,  
Antonarakis,S.E., Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.  
TITLE Positional cloning of the APECD gene  
JOURNAL Nat. Genet. 17 (4), 393-398 (1997)  
MEDLINE 98061086  
PUBMED 9398839

REFERENCE  
AUTHORS 2 (baaes 1 to 1545)  
TITLE Direct Submision  
JOURNAL Submitted (16-AUG-1997) Nobuyoshi Shimizu, Keio University School  
of Medicine, Department of Molecular Biology, 35 Shinanomachi,  
Shinjuku-Ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp,  
Tel:03-3351-2370, Fax:03-3351-2370)  
FEATURES  
source location/Qualifiers  
1..1545  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolate="Caucasian"  
/db\_xref="taxon:9606"  
/chromosome="21"



TITLE Kite, such as nucleic acid arrays, comprising a majority of humenexons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 17797 06-SEP-2002;  
PE Corporation (NY) (US)  
FEATURES Location/Qualifiers  
source 1.1545  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 99.7%; Score 1540.2; DB 6; Length 1545;  
Best Local Similarity 99.8%; Pred. No. 8.7e-290;  
Matches 1542; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AGAGAAAGTGAAGTCTTCTCAGGCTTTAAAGCATGGGTTGTCTCAGGCTGTAACCG 60  
Db 1 AGAGAAAGTGAAGTCTTCTCAGGCTTTAAAGCATGGGTTGTCTCAGGCTGTAACCG 60  
QY 61 CTGCTCTCAGTGGGCGCGGTGGGCGCGGCGCGCGCGCGCGCTTATACCAAGAGTCAAG 120  
Db 61 CTGCTCTCAGTGGGCGCGGTGGGCGCGGCGCGCGCGCGCTTATACCAAGAGTCAAG 120  
QY 121 ATCCACTGGGAATGCCATGCTCATCTTTCGTCCCAAGCATGGTTCTTAATGGGGTAA 180  
Db 121 ATCCACTGGGAATGCCATGCTCATCTTTCGTCCCAAGCATGGTTCTTAATGGGGTAA 180  
QY 181 GCAGGTGGGAGAGACCTCCCTGGGCGCTGGCCCACTGCTGTGAGAAAGGTTGAT 240  
Db 181 GCAGGTGGGAGAGACCTCCCTGGGCGCTGGCCCACTGCTGTGAGAAAGGTTGAT 240  
QY 241 GGTGGTGTACAGTTCCGGGGCGCGCGGCGCGGCGCGGCGCGGCGGCGGCGGCGG 300  
Db 241 GGTGGTGTACAGTTCCGGGGCGCGCGGCGCGGCGCGGCGCGGCGGCGGCGGCGG 300  
QY 301 TCCCATAGAGATGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
Db 301 TCCCATAGAGATGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 GTGAGGGGCAAGATTTCAGGCGCTGGAGCATGGGAGCGGAGAGAGAGAGAGAGAG 420  
Db 361 GTGAGGGGCAAGATTTCAGGCGCTGGAGCATGGGAGCGGAGAGAGAGAGAGAGAG 420  
QY 421 GTACCCAGAGATGCTGCGGGGAGCTGTTTGGGAAAGAGAGAGAGAGAGAGAGAG 480  
Db 421 GTACCCAGAGATGCTGCGGGGAGCTGTTTGGGAAAGAGAGAGAGAGAGAGAGAG 480  
QY 481 TGCACCCAGAGAGTCTGATGAGGCGTCTTTCCTGCTGCTCAGAAAGATGAGAGAG 540  
Db 481 TGCACCCAGAGAGTCTGATGAGGCGTCTTTCCTGCTGCTCAGAAAGATGAGAGAG 540  
QY 541 GTGCGCTGTGTGCGGAGCGGCGGAGAGTCACTGCTGTGAGAGAGAGAGAGAGAG 600  
Db 541 GTGCGCTGTGTGCGGAGCGGCGGAGAGTCACTGCTGTGAGAGAGAGAGAGAGAG 600  
QY 601 ACTGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Db 601 ACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
QY 661 GGTGCTGAGAGCAAGTTCAGAGAGTCAAGCCCGGAGAGAGAGAGAGAGAGAGAG 720  
Db 661 GGTGCTGAGAGCAAGTTCAGAGAGTCAAGCCCGGAGAGAGAGAGAGAGAGAGAG 720  
QY 721 AGCCACCTGAG 780  
Db 721 AGCCACCTGAG 780  
QY 781 GTCCACCTGAG 840  
Db 781 GTCCACCTGAG 840  
QY 841 CTCGCGCTTCTGAG 900

Db 841 CTCGCGCTTCTGAG 900  
QY 901 GTGAGGCTCTGAG 960  
Db 901 GTGAGGCTCTGAG 960  
QY 961 ATGTACAG 1020  
Db 961 ATGTACAG 1020  
QY 1021 TCCAG 1080  
Db 1021 TCCAG 1080  
QY 1081 TGACCCAG 1140  
Db 1081 TGACCCAG 1140  
QY 1141 CTGCAAG 1200  
Db 1141 CTGCAAG 1200  
QY 1201 TTCTGAG 1260  
Db 1201 TTCTGAG 1260  
QY 1261 CGGCG 1320  
Db 1261 CGGCG 1320  
QY 1321 TGTGAG 1380  
Db 1321 TGTGAG 1380  
QY 1381 GACAGGCGAG 1440  
Db 1381 GACAGGCGAG 1440  
QY 1441 GGCATCATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
Db 1441 GGCATCATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 GCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
Db 1501 GCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
RESULT 3  
AB006685 1463 bp mRNA linear Pri 13-DEC-1997  
LOCUS AB006685  
DEFINITION Homo sapiens APECED mRNA for AIRE-3, complete cds.  
ACCESSION AB006685.1 GI:2696622  
VERSION AB006685.1  
KEYWORDS APECED; AIRE-3.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Nagamine, K., Peterson, P., Scott, H., S., Kudoh, J., Minoshima, S.,  
1 (sites)  
Heino, M., Krohn, K., J. E., Laijoi, M., D., Mullis, P., E.,  
Antonarakis, S. E., Kawasaki, K., Asakawa, S., Ito, F. and Shimizu, N.  
TITLE Positional cloning of the APECED gene  
JOURNAL Nat. Genet. 17 (4), 393-398 (1997)  
MEDLINE 9398839  
REFERENCE 2 (bases 1 to 1463)  
AUTHORS Shimizu, N.  
TITLE Direct Submission  
JOURNAL Submitted (16-AUG-1997) Nobuyoshi Shimizu, Keio University School  
of Medicine, Department of Molecular Biology, 35 Shinanomachi,  
Shinjuku-ku, Tokyo 160, Japan (E-mail: shimizu@mb.med.keio.ac.jp).



FEATURES  
Tel:03-3351-2370, Fax:03-3351-2370)

## source

Location/Qualifiers  
1. 1463

/organism="Homo sapiens"

/mol\_type="mRNA"

/isolate="Caucasian"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/sex="male"

/issue\_type="thymus"

/dev\_stage="3-yr-old"

1. 1463

/gene="APECD"

237. 1001

/gene="APECD"

/note="autoimmune regulator-3"

/codon\_start=1

/product="AIRE-3"

/protein\_id="BA023993.1"

/db\_xref="GI:2696623"

/translation="MWLVSSGAPGTQAPARNVFPPIGMAPGVCMRPDGMGTGOG  
RISPGSMAGORLSSGTORCCMSGFGKEVALRVLHPSPYCMGVSTLCOINDEEC  
AVCRDGGELICDGPAPHLACISPLREIPSTWRGSSCLQATQVOPRAEERBP  
QEPVETPLPBCIRASGEPRCQGTPTPTCTPTCTVWLVKSRVTLVRAAGCAEMVRT  
CCGVLTALPSTGATSTQPAPEPGACADPAQET"

1449. 1454

/gene="APECD"

polya\_signal

## ORIGIN

Query Match 88.5%; Score 1367.8; DB 9; Length 1463;

Best Local Similarity 94.6%; Pred. No. 3.3e-256;

Matches 1461; Conservative 0; Mismatches 2; Indels 82; Gaps 1;

1 AGAGAAAGTAGAGTCTTCTCAGGCTCTTAAGAGCATGCGTTGTCTCAGGCTGTACCG 60  
1 AAGAGAAAGTAGAGTCTTCTCAGGCTCTTAAGAGCATGCGTTGTCTCAGGCTGTACCG 60  
61 GTGCTCTCAGCTGGGCGCGGTGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
61 CTGCTCTCAGCTGGGCGCGGTGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
121 ATCCACTGGGAATGCATGCTCATCTTTGTCTCCCGAGCATGCTTTCTTAATGGGGTAA 180  
121 ATCCACTGGGAATGCATGCTCATCTTTGTCTCCCGAGCATGCTTTCTTAATGGGGTAA 180  
181 GCAGGTCGGGAGAGACTCTCTGAGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
181 GCAGGTCGGGAGAGACTCTCTGAGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
241 GGTGCTGTACAGTTCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
241 GGTGCTGTACAGTTCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
301 TCCCAATAGGAGATGGGCG 360  
301 TCCCAATAGGAGATGGGCG 360  
361 GTACAGGCGAGAAATTTACAGGCGCGCGAGCATGGGAGAGAGAGAGAGAGAGAGAGAG 420  
361 GTACAGGCGAGAAATTTACAGGCGCGCGAGCATGGGAGAGAGAGAGAGAGAGAGAGAG 420  
421 GTACAGGCGAGATGCTGCTGGGAGAGCTGTTTGGAGAGAGAGAGAGAGAGAGAGAG 480  
421 GTACAGGCGAGATGCTGCTGGGAGAGCTGTTTGGAGAGAGAGAGAGAGAGAGAGAG 480  
481 TGAACCCCGAGCGAGCTGAGTGGGAGCTGCTTGGCGAGAGAGAGAGAGAGAGAGAG 540  
481 TGAACCCCGAGCGAGCTGAGTGGGAGCTGCTTGGCGAGAGAGAGAGAGAGAGAGAG 540  
541 GTGCGCTGTGTGGGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
541 GTGCGCTGTGTGGGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

601 ACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
601 ACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
661 GCTGCTGAG 720  
661 GCTGCTGAG 720  
721 AGCAGCCGCTGAG 780  
721 AGCAGCCGCTGAG 780  
781 GTCCAGCTGGGAG 840  
781 GTCCAGCTGGGAG 840  
841 CTCGCGCTTCTGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
841 CTCGCGCTTCTGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
901 GTGAGGCTCTGAG 960  
901 GTGAGGCTCTGAG 960  
961 ATGATAG 1020  
961 ATGATAG 1020  
1021 TCCAGCGCGAG 1080  
1021 TCCAGCGCGAG 1080  
939 TCCAGCGCGAG 998  
939 TCCAGCGCGAG 998  
1081 TGAACCCGAG 1140  
1081 TGAACCCGAG 1140  
999 TGAACCCGAG 1058  
999 TGAACCCGAG 1058  
1141 CTGCGAG 1200  
1141 CTGCGAG 1200  
1059 CTGCGAG 1118  
1059 CTGCGAG 1118  
1201 TTCTGAG 1260  
1201 TTCTGAG 1260  
1119 TTCTGAG 1178  
1119 TTCTGAG 1178  
1261 CGGCG 1320  
1261 CGGCG 1320  
1179 CGGCG 1238  
1179 CGGCG 1238  
1321 TGTGAG 1380  
1321 TGTGAG 1380  
1239 TGTGAG 1298  
1239 TGTGAG 1298  
1381 GACAGCGAG 1440  
1381 GACAGCGAG 1440  
1299 GACAGCGAG 1358  
1299 GACAGCGAG 1358  
1441 GCCATGATGCTGAG 1500  
1441 GCCATGATGCTGAG 1500  
1359 GCCATGATGCTGAG 1418  
1359 GCCATGATGCTGAG 1418  
1501 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
1501 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545  
1419 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1463  
1419 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1463

## RESULT 4

AB006682 2027 bp mRNA linear PRI 13-DEC-1997

LOCUS Homo sapiens APECD mRNA for AIRE-1, complete cds.

DEFINITION AB006682  
AB006682  
AB006682.1 GI:2696614  
KEYWORDS APECD; AIRE-1.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S., Heino,M., Krohn,K.J.E., Lalioti,M.D., Mullis,P.E., Antonarakis,S.E., Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N. Anticardiolin, a novel member of the APECD gene family. Nat. Genet. 17 (4), 393-398 (1997)  
 JOURNAL Nat. Genet. 17 (4), 393-398 (1997)  
 MEDLINE 98061086  
 PUBMED 9398839  
 REFERENCE 2 (bases 1 to 2027)  
 AUTHORS Shimizu,N.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-1997) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)  
 FEATURES  
 source  
 1. 2027  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /isolate="Caucasian"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21q22.3"  
 /sex="male"  
 /tissue\_type="thymus"  
 /dev\_stage="3-yr-old"  
 1. 2027  
 /gene="APECED"  
 128. 1765  
 /note="autoimmune regulator-1"  
 /product="AIRE-1"  
 /protein\_id="BA23988.1"  
 /db\_xref="GI:2696615"  
 /translation="MATDAALRLRLALHRTETAVAVDSAPFLHIALADHDVPEKQ  
 ETLHKEKSGCPDPAFLALSWLITDSTAILDPKRLVLPDYNERYGLQPLIDSPK  
 DVLDQPRGRKRPVPAKLVPPRLPTRKASBEKRAAPALTRGTASPSQLKA  
 KPPKKESSAEDQRLPLNGIQTMSASVORAVASGSDVPAAGAVEGLIQVFESG  
 GSKKCIQVGEFTYTPSPKEDSGSKKASVSPKPLVAGAKOGAPGSGEIRLQO  
 GSVAPALPSPDQLHOKMEDCAVCRDGBELICCGCPRAFLACLSPKIRISGT  
 WRCSSCIQATVQVOPRAEPRBPVTPPLPGIRSGAEVREGPRPGLRMDPTL  
 VYKHLPAPPSAPPLPGLDSSALHPLICVPEBGGQNALPARGCVCDGTDLACTICA  
 PALHMDLSLSLSEHTFDGILQWALQSMARPAAPFS"  
 2013. 2018  
 /gene="APECED"  
 ORIGIN  
 polyA\_signal  
 Query Match 66.3%; Score 1023.6; DB 9; Length 2027;  
 Best Local Similarity 99.6%; Pred. No. 3.3e-189;  
 Matches 1026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 516 CTGTGCAGAGAATAGAGACGAGTGCCGTGTGCGGAGCGCGGAGCTCATCTGC 575  
 DB 998 CTCACACAGAGATAGAGACGAGTGCCGTGTGCGGAGCGCGGAGCTCATCTGC 1057  
 QY 576 TGTGACGGCTGCCCTCGGAGCTTCCACCTGAGCTGCTGCCCTCGGAGCTGAGANT 635  
 DB 1058 TGTGACGGCTGCCCTCGGAGCTTCCACCTGAGCTGCTGCCCTCGGAGCTGAGANT 1117  
 QY 636 CCCAGTGGAGCTGAGAGTGCTCAAGCTGCTGACAGCAACAGTCCAGAGGTGACAGCC 695  
 DB 1118 CCCAGTGGAGCTGAGAGTGCTCAAGCTGCTGACAGCAACAGTCCAGAGGTGACAGCC 1177  
 QY 696 CGGAGAGAGAGCGCGGAGCGAGAGCAACCGGTGAGAGCGCGGAGCGCGGAGCGCTT 755  
 DB 1178 CGGAGAGAGAGCGCGGAGCGAGAGCAACCGGTGAGAGCGCGGAGCGCGGAGCGCTT 1237

QY 756 AGATCGGCGGAGAGAGAGTAAAGATCCACCTGAGGAAACCCCTAGCCGAGTACAGCAG 815  
 DB 1238 AGATCGGCGGAGAGAGAGTAAAGATCCACCTGAGGAAACCCCTAGCCGAGTACAGCAG 1297  
 QY 816 ACTTTGTCTACAGACACTGCGGCTCGCTTCTGACGCCCCGCTGCAAGTCTGAC 875  
 DB 1298 ACTTTGTCTACAGACACTGCGGCTCGCTTCTGACGCCCCGCTGCAAGTCTGAC 1357  
 QY 876 TCCTGCGGCTGACACCCCTACTGTTGTGGGTCCGAGAGGTGACAGAACCGGCTCT 935  
 DB 1358 TCCTGCGGCTGACACCCCTACTGTTGTGGGTCCGAGAGGTGACAGAACCGGCTCT 1417  
 QY 936 GGTGCGCTTGGGAGGTGTCGAGATGTAACGAGTCTGCTGCTGCTGCTGCTGCTG 995  
 DB 1418 GGTGCGCTTGGGAGGTGTCGAGATGTAACGAGTCTGCTGCTGCTGCTGCTGCTG 1477  
 QY 996 GCTGCTTCTACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1055  
 DB 1478 GCTGCTTCTACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1537  
 QY 1056 CGCTGAGATCTGCTGAGAGACGTAACCCGAGCCCTGAGAGGAGGTGCTGAGCCGC 1115  
 DB 1538 CGCTGAGATCTGCTGAGAGACGTAACCCGAGCCCTGAGAGGAGGTGCTGAGCCGC 1597  
 QY 1116 AGCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1175  
 DB 1598 AGCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1657  
 QY 1176 CTGACAGAGATGACTTGAATGCTCTTCTGAGCGAGACACCTTGTGATGCTCTG 1235  
 DB 1658 CTGACAGAGATGACTTGAATGCTCTTCTGAGCGAGACACCTTGTGATGCTCTG 1717  
 QY 1236 TGGGCGATCAGAGATGAGCGGCTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1295  
 DB 1718 TGGGCGATCAGAGATGAGCGGCGGCTGCGAGGCGGCGGCGGCGGCGGCGGCGG 1777  
 QY 1296 GGAACATGAGCTGATGAGAGAGTGTGAGAGAGACCTTCTCTCTCTCTCTCTG 1355  
 DB 1778 GGAACATGAGCTGATGAGAGAGTGTGAGAGAGACCTTCTCTCTCTCTCTCTG 1837  
 QY 1356 GCGGCGGCTGATGATGAGAGAGGAGACGCGCACCTTCTCTCTCTCTCTCTCTG 1415  
 DB 1838 GCGGCGGCTGATGATGAGAGAGGAGACGCGCACCTTCTCTCTCTCTCTCTCTG 1897  
 QY 1416 CAGCTCTGTTCTGAGAGACAGAGCATGATGCTGAGAAATTAACCTGAGCCAC 1475  
 DB 1898 CAGCTCTGTTCTGAGAGACAGAGCATGATGCTGAGAAATTAACCTGAGCCAC 1957  
 QY 1476 TTCTCTACTCTGAAAGTCCCGGAGCTCTCTCTGCTGAGCTACTATAAATATTA 1535  
 DB 1958 TTCTCTACTCTGAAAGTCCCGGAGCTCTCTCTGCTGAGCTACTATAAATATTA 2017  
 QY 1536 AAATTAAGCTG 1545  
 DB 2018 AAATTAAGCTG 2027  
 RESULT 5  
 A97284 2245 bp DNA linear PAT 26-JAN-2000  
 LOCUS A97284  
 DEFINITION Sequence 1 from Patent WO918197.  
 ACCESSION A97284  
 VERSION A97284.1 GI:6780668  
 KEYWORDS  
 SOURCE unidentifed  
 ORGANISM unidentifed  
 REFERENCE 1 (bases 1 to 2245)  
 AUTHORS Yasuo,M. and Lehar,H.  
 TITLE NUCLEIC ACID MOLECULE ENCODING A (POLY)PEPTIDE CO-SEGREGATING IN MUTATED FORM WITH AUTOIMMUNE POLYENDOCRINOPATHY CANDIDIASIS  
 ECTODERMAL DYSTROPHY (APECED)  
 JOURNAL Patent: WO 918197-A.1 15-Apr-1999;







Query Match	Best Local Similarity	Score	DB	Length
29.9%	97.7%	462.4	20000	
469; Conservative	0; Mismatches	11; Indels	0; Gaps	0;
183	AGGTCGGAGAGACCTCCCTGGAGCTGAGCCCACTGCTCTGTGAGGAAGGTTTCATGTGG	242		
9453	AGGTCGGAGAGACCTCCCTGGAGCTGAGCCCACTGCTCTGTGAGGAAGGTTTCATGTGG	9512		
243	TTGGTGTACAGTTCCGGGGGCCCTTGGAACGACGAGCTGCAAGAAACGGGGTTTCTTC	302		
9513	TTGGTGTACAGTTCCGGGGGCCCTTGGAACGACGAGCTGCAAGAAACGGGGTTTCTTC	9572		
303	CCAAATAGGGAAGGCGCGGGGGGTCTCTGTTCGAGACCAAGATTGATGGGGAACAGGTGGT	362		
9573	CCAAATAGGGAAGGCGCGGGGGGTCTCTGTTCGAGACCAAGATTGATGGGGAACAGGTGGT	9632		
363	CAGGCGAATTTTCAGGCTTGGACGATGGGAGCGGGGAGAGACTGGGGAGTTTCAGGT	422		

Db	9633	CAGGGCAGAAATTTCAGAGCCCTCGCAGCATGAGGACAGGACAGAGACTGCGGGAGTTCAAGT	96932
Qy	423	ACCCAGAGATGCTCTGCGGGAGAGCTGTTTTGGAGAGAGTGCTCTCAGAGAGGTGCTG	482
Db	9693	ACCCAGAGATGCTGCTCGGGGAGAGCTGTTTTGGAGAGAGGTGCTTCAGAGAGGTGCTG	97522
Qy	483	CACCCAGAGCCAGCTGCATGAGGAGGCGTCTTTGCGCTGAGCCAGAAAGTAAGACAGATGT	542
Db	9753	CACCCAGAGCCAGCTGTCATGAGGAGGCGTCTTTGCGCTGAGCCAGAAATGAGAGCAGATGT	98122
Qy	543	GCCGCTGTGTGCGGACGGCGGGAGCTCATCTGTGTGACAGGCTGCGCTCGGGCTTTCCAC	602
Db	9813	GCCGTGTGTGCGGACGGCGGGAGCTCATCTGTGTGACAGGCTGCGCTCGGGCTTTCCAC	98722
Qy	603	CTGCGCCGTCCTGTCCTCCCTCGCTCCCGGAGATATCCCAAGTGGAGATGAGAGTGTCTCAGC	662
Db	9873	CTGCGCCGTCCTGTCCTCCCTCGCTCCCGGAGATCCCAAGTGGAGCTGACCTCTGCGAGC	99322
RESULT 9			
LOCUS	HSNJ9610	36284 bp	DNA linear PRI 01-SEP-2000
DEFINITION	Homo sapiens AIRE gene.		
ACCESSION	AJ009610		
VERSION	AJ009610.1	GI:3392939	
KEYWORDS	Aire gene.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Bukatyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 Maltounen, J., Björsev, P., Perheentupa, J., Horelli-Kuitman, N., Palot, A., Peltonen, L., Lee, Y.S., Francis, F., Hennig, S., Thiel, C., Lehrach, H. and Yaspo, M.L.		
TITLE	An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PRD-type zinc finger domains		
JOURNAL	Nat. Genet. 17, 399-403 (1997)		
MEDLINE	2 (bases 1 to 36284)		
REFERENCE	Lee, Y.S., Francis, F., Hennig, S., Thiel, C., Reinhard, R., Lehrach, H. and Yaspo, M.L.		
AUTHORS	Direct Submission		
TITLE	Submitted (17-JUL-1998) Sreiffen Hennig, MPIWG, Abt. Lehrach, Max Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany		
JOURNAL	Related cDNA sequence: Z97990.		
COMMENT	Location/Qualifiers		
FEATURES	1..36284		
SOURCE	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="21"		
	/map="q22.3"		
	/clone_id="Lawrence Livermore Cosmid 22G11"		
	199..323		
	/note="11MC3"		
repeat_region	complement(779..10930)		
repeat_region	/note="AluY"		
repeat_region	complement(1101..1198)		
repeat_region	/note="AluSc"		
repeat_region	complement(11245..1533)		
repeat_region	/note="AluSg"		
repeat_region	complement(1537..1678)		
repeat_region	/note="AluY"		
repeat_region	complement(1668..1823)		
repeat_region	/note="AluSc"		
repeat_region	complement(1824..2106)		
repeat_region	/note="AluSg"		
repeat_region	complement(2112..2196)		
repeat_region	/note="AluSg/X"		
repeat_region	complement(2208..2234)		
repeat_region	/note="(GAAA)n"		
repeat_region	complement(2235..2532)		

repeat_region	/note="AluX"	complement(3109. .3382)	gene	22529. .34854
repeat_region	/note="AluY"	complement(5332. .5634)	exon	/gene="AIRE"
repeat_region	/note="AluSg"	complement(5722. .5820)		22529. .22779
repeat_region	/note="(GAAA)n"	6217. .6515		/gene="AIRE"
repeat_region	/note="AluSg"	6531. .6761	CDS	/number=1
repeat_region	/note="AluO"	complement(6805. .7106)		join(22648. .22779,23198. .23372,23619. .23774,24158. .24232,
repeat_region	/note="AluSp"	complement(7107. .7409)		24986. .25099,26298. .26443,26629. .26709,27336. .27851,
repeat_region	/note="AluSp"	7445. .7729		28943. .29042,29633. .29815,30428. .30459,31040. .31142,
repeat_region	/note="AluSg"	complement(7757. .7798)		33022. .33084,34291. .34362)
repeat_region	/note="(GAAA)n"	complement(7799. .8101)		/gene="AIRE"
repeat_region	/note="AluSx"	10206. .10439		/codon start=1
repeat_region	/note="LI"	10443. .10742		/protein_id="CAA08759.1"
repeat_region	/note="AluSx"	complement(10918. .11215)	exon	/db_xref="Swiss-Prot:O43918"
repeat_region	/note="AluSx"	complement(11501. .11878)	exon	/db_xref="Swiss-Prot:O43918"
repeat_region	/note="MSTB"	complement(11993. .12079)		/translation="MATDAALRLRLRLHRTETIAVDSAPFLHLALADHDVVEBDFQ
repeat_region	/note="MSTB"	12193. .12880	exon	ETLHLKEKGCPOAFHALISMLTODSTALIDFMRYLFQDYNLERGRLOPILDSFPK
repeat_region	/note="LI"	complement(12887. .13178)	exon	DVDLSOPRRGRKRPVAPKALVPPRLPTFRKASEERAAAPALPGRGAFSPGLKA
repeat_region	/note="AluSc"	13289. .13582		KPKKDESAEQOQLPLNGIOTMSASVORAVAMSGVPGARGAVEGILIOVPESG
repeat_region	/note="AluSx"	13585. .13738		GSKKCIQGGERTPTSKTEDSGSKAKRSSGPKPLVRAKAGQAPOGGEARLQO
repeat_region	/note="LI"	complement(13751. .14060)		GSVPAPLALPSDPOLHQKNEDECAVCRDGGELICDGCPRAFHLCLSPLELIPBGT
repeat_region	/note="LI"	complement(14061. .14357)		WRSSCLQATVQEVQPRAEPRPOEPVETPLPGILRSAGEVRGPPGRLGMDTTL
repeat_region	/note="AluSg"	complement(14360. .14720)	exon	VYKHLPAEPSAAPLPGLDSALHPILCVGPEGOQMLAPARGVMTGLRCRSCSDVT
repeat_region	/note="LI"	complement(14722. .15018)		PAAPPS"
repeat_region	/note="AluY"	complement(15019. .15394)	exon	23198. .23372
repeat_region	/note="LI"	complement(15398. .15700)		/gene="AIRE"
repeat_region	/note="AluSg"	complement(15704. .15864)	repeat_region	/number=2
repeat_region	/note="LI"	15865. .16182	repeat_region	23619. .23774
repeat_region	/note="LI"	16726. .16931	exon	/gene="AIRE"
repeat_region	/note="AluSg/x"	complement(16933. .17084)		/number=3
repeat_region	/note="MLTIA1"	complement(17090. .17375)		24158. .24232
repeat_region	/note="AluSx"	complement(17379. .17530)	exon	/gene="AIRE"
repeat_region	/note="MLTIA1"	17663. .17959		/number=4
repeat_region	/note="AluY"	18738. .18920	exon	24986. .25099
repeat_region	/note="MIR"	complement(19498. .20418)		/gene="AIRE"
repeat_region	/note="LTR5"	complement(21877. .21968)	exon	29633. .29815
repeat_region	/note="MIR"	complement(22524. .22642)	repeat_region	/number=10
repeat_region	/note="GC-rich"		repeat_region	/gene="AIRE"
			repeat_region	/number=11
			repeat_region	complement(30872. .30983)
			exon	/note="POLY G"
			exon	31040. .31152
			exon	/gene="AIRE"
			exon	/number=12
			exon	33022. .33084
			exon	/gene="AIRE"
			exon	/number=13
			exon	34291. .34854
			exon	/gene="AIRE"
			exon	/number=14
			exon	34600. .34787
			exon	/note="AluSg"
			exon	complement(34793. .35417)



```
repeat_region complement(35423..35747)
/clone_lib="KU21-D library"
Query Match 29.9%; Score 462.4; DB 9; Length 42133;
Best Local Similarity 97.7%; Pred. No. 3.9e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 242
|
27394 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 27453
|
243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
|
27454 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 27513
|
303 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 362
|
27514 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 27573
|
363 CAGGGCAGAAATTTCAAGCCCTGGACATGGAGAGAGGAGAGAGACTGGGGAGTTCAAGT 422
|
27574 CAGGGCAGAAATTTCAAGCCCTGGACATGGAGAGAGGAGAGAGACTGGGGAGTTCAAGT 27633
|
423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGAGTCTG 482
|
27634 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGAGTCTG 27693
|
483 CACCCAGAGCTGCTGCAATGGGCGTCTCTTGGCTTGCAGAAAGATGAGACAGATGT 542
|
27694 CACCCAGAGCTGCTGCAATGGGCGTCTCTTGGCTTGCAGAAAGATGAGACAGATGT 27753
|
543 GCCGTGTGCGGAGCGGCGGGAGCTCATCTGTGTGAGCGGCTGGCCCTGGGGCTTCCAC 602
|
27754 GCCGTGTGCGGAGCGGCGGGAGCTCATCTGTGTGAGCGGCTGGCCCTGGGGCTTCCAC 27813
|
603 CTGGCTGTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGGACCTGAGAGTGTCCAC 662
|
27814 CTGGCTGTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGGACCTGAGAGTGTCCAC 27873
|

RESULT 10
AP001060 42133 bp DNA linear PRI 14-APR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21, clone:D4G11, MX1-D21S171
DEFINITION Homo sapiens genomic DNA, chromosome 21, clone:D4G11, MX1-D21S171
ACCESSION AP001060
VERSION AP001060.1 GI:6693610
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 42133)
Shimizu,N., Kudoh,J. and Shibuya,K.
Homo sapiens genomic DNA, chromosome 21, clone D4G11, MX1-D21S171
TITLE
REFERENCE
AUTHORS
JOURNAL
PUBLISHED Only in Database (2000)
2 (bases 1 to 42133)
Shimizu,N., Kudoh,J. and Shibuya,K.
Direct Submision
Submitted (12-JAN-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES
source
1..42133
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/clone="D4G11"
/cell_line="GM130B"
/cell_type="B-lymphoblastoid"
```

```
ORIGIN
/clone_lib="KU21-D library"
Query Match 29.9%; Score 462.4; DB 9; Length 42133;
Best Local Similarity 97.7%; Pred. No. 3.8e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 242
|
30870 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 30929
|
243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
|
30930 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 30989
|
303 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 362
|
30990 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 31049
|
363 CAGGGCAGAAATTTCAAGCCCTGGACATGGAGAGAGGAGAGAGACTGGGGAGTTCAAGT 422
|
31050 CAGGGCAGAAATTTCAAGCCCTGGACATGGAGAGAGGAGAGAGACTGGGGAGTTCAAGT 31109
|
423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGAGTCTG 482
|
31110 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGAGTCTG 31169
|
483 CACCCAGAGCTGCTGCAATGGGCGTCTTGGCTTGCAGAAAGATGAGACAGATGT 542
|
31170 CACCCAGAGCTGCTGCAATGGGCGTCTTGGCTTGCAGAAAGATGAGACAGATGT 31229
|
543 GCCGTGTGCGGAGCGGCGGGAGCTCATCTGTGTGAGCGGCTGGCCCTGGGGCTTCCAC 602
|
31230 GCCGTGTGCGGAGCGGCGGGAGCTCATCTGTGTGAGCGGCTGGCCCTGGGGCTTCCAC 31289
|
603 CTGGCTGTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGGACCTGAGAGTGTCCAC 662
|
31290 CTGGCTGTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGGACCTGAGAGTGTCCAC 31349
|

RESULT 11
AC003656_5
WPCOMMENT
Sequence split into 8 fragments LOCUS AC003656 Accession AC003656
Fragment Name Begin End
AC003656_1 100001 210000
AC003656_2 200001 310000
AC003656_3 300001 410000
AC003656_4 400001 510000
AC003656_5 500001 610000
AC003656_6 600001 710000
AC003656_7 700001 753000
Continuation (6 of 8) of AC003656 from base 500001 (AC003656 Homo sapiens clone PI C124)

Query Match 29.9%; Score 462.4; DB 2; Length 110000;
Best Local Similarity 97.7%; Pred. No. 3.3e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 242
|
52370 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGGCCCTGTGAGAGAGGTTCAATGTG 52429
|
243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
|
52430 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 52489
|
303 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 362
|
52490 CCAATTAGGAGTGGCCCCCGGGGGGTGTCTGTTCAGAACCAAGATGGGGAAACAGGTGT 52549
|
363 CAGGGCAGAAATTTCAAGCCCTGGACATGGAGAGAGGAGAGAGACTGGGGAGTTCAAGT 422
|
```

Db	52550	CAGGGCAGAATTTTCAAGCCCTTCGCGACAGATGAGGACAGGGCAGAGACTGTGGGAAGTTTCAGGT	52603
Oy	423	ACCACAGAGATGCTCTCTGGGGAAGCTGTTTTGGAGAGAGTGCTCTCAGAGAGGTCGTG	482
Db	52610	ACCACAGAGATGCTCTCTGGGGAAGCTGTTTTGGAGAGAGTGCTCTCAGAGAGGTCGTG	52663
Oy	483	CACCCAGGCCAGCTCTGCATGGGGGCTCTTGGCTGTGCCAAGAAATAGAGCGAGTGT	542
Db	52670	CACCCAGGCCAGCTCTGCATGGGGGCTCTTGGCTGTGCCAAGAAATAGAGCGAGTGT	52729
Oy	543	GCCGTGTGTGCGGAGCGCGGGGAGCTCATCTGCTGTGTAACGGTGCCTCGGGGCTTCCAC	602
Db	52730	GCCGTGTGTGCGGAGCGCGGGGAGCTCATCTGCTGTGTAACGGTGCCTCGGGGCTTCCAC	52785
Oy	603	CTGGCCTGTGCTTCTCCCTCCGCTTCGGGAGATCCCAAGTGGACCTTGAAGTGTCTCAC	662
Db	52790	CTGGCCTGTGCTTCTCCCTCCGCTTCGGGAGATCCCAAGTGGACCTTGAAGTGTCTCAC	52849
RESULT_12			
AP001754			
LOCUS		340000 bp	DNA linear PRI 06-JUN-2003
DEFINITION	Homo sapiens genomic DNA, chromosome 21q, section 98/105.		
ACCESSION	AP001754 AL163299 BA000005		
VERSION	AP001754.1 GI:7768775		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Tauden,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Saeki,K., Nagamine,K., Mitsuyma,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordieck,G., Hontelaicher,K., Barand,P., Schafte,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecher,H., Ramser,J., Beck,A., Klages,S., Henning,S., Riesemann,L., Dsgand,E., Wehmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L. The DNA sequence of human chromosome 21 Nature 405 (6784), 311-319 (2000) 20289799 10830953 2 (bases 1 to 340000)		
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

```

* e.mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscf-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
160-8582, Japan,
* e.mail: nshimizudmb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@bf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
Inestrassse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163299: Submitted (10-Apr-2000).
Location/Qualifiers
1. 34000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
<1. 17637
/map="21q22.3"
/clone_id="KUD4G11, 5' partial"
/clone_lib="KU21-D Cosmid library"
/note="Accession No. AP001060"
16296. 32142
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_id="KUD6A4"
/clone_lib="KU21-D Cosmid library"
/note="Accession No. AP001061"
25925. 62855
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_id="Q5B10"
/clone_lib="L121NCO2-Q Cosmid library"
/note="Accession No. AP001062"
56250. 101561
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_id="KUD99F9"
/clone_lib="KU21-D Cosmid library"
/note="Accession No. AP001063"
97304. 135091
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_id="KUD107C4"
/clone_lib="KU21-D Cosmid library"

```

source /note="Accession No. AP001064"  
109897. .225790  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KB68A7"  
/clone\_1ib="Keio BAC library"  
/note="Accession No. AP001065"  
212299. .251501  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KUD1H9"  
/clone\_1ib="KUD1-D Cosmid library"  
/note="Accession No. AP001066"  
230165. .>340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="KB139C7, 3' partial"  
/clone\_1ib="Keio BAC library"  
/note="Accession No. AP001067"  
280806. .>340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="P314N7, 3' partial"  
/clone\_1ib="RPC11, 3-5 PAC library"  
/note="Accession No. AJ011930"  
1508. .13840  
/gene="AIRE"  
join(1508. .1759,2178. .2352,2599. .2754,3138. .3212,  
3966. .4079,5278. .5423,5609. .5689,6716. .6831,7923. .8022,  
8614. .8796,9410. .9531,10022. .10124,12004. .12066,  
13277. .13840)  
/gene="AIRE"  
/product="autoimmune regulator (APECED protein)"  
/note="Accession No. Z97990"  
1508. .1759  
/gene="AIRE"  
/number=1  
1508. .1627  
/gene="AIRE"  
join(1628. .1759,2178. .2352,2599. .2754,3138. .3212,  
3966. .4079,5278. .5423,5609. .5689,6716. .6831,7923. .8022,  
8614. .8796,9410. .9531,10022. .10124,12004. .12066,  
13277. .13348)  
/gene="AIRE"  
/note="Accession No. Z97990"  
/codon\_start=1  
/product="autoimmune regulator (APECED protein)"  
/protein\_id="BAA95560.1"  
/db\_xref="GI:7768776"  
translation="MATDALRLRLRLHRTETAVVDSAPFLHLALADHDVPEBDFQ  
ETLHLKEKCCQAPALSWLTDSTALIDPFWLFDVNLERYGRLOPLIDSPK  
VDLSQPRKGRPAVPKALVPPRLPTGRKASBARAPALTPRGTSQSLKA  
KPKKPSABOOLPLINGIOTMSAVORAVAMSSGVPRGAGAVEGILLIOVESG  
GSKKCIQVGEERTPSKPSDSSGKKARSSGPPPLVAKAKGAPGEGARLQGO  
GSVPALPSPDOLQKNEDECAVCRDGLICCDGCPRAFLACLSPLELISGTT  
WCSVCLQATVQVQPRABEPPOBPVETPLPGRLSAGEVETGPPGEPPLAGMTTL  
VKHLPAFSAAPLPLGLSSALPLLCVPEQONLAPARCGVCGDGDVLRCTHA  
AAFWHCHPACTSRPGTGLRCRSCSGDVTAPVGVGLAPSPARLAPAPAKODTNSHE  
PALHRDLESLSEHTFQIILQWATQSMARPAAPPS"  
2178. .2352  
/gene="AIRE"

5'UTR  
CDS  
exon

exon /number=2  
2599. .2754  
Query Match 29.9%; Score 462.4; DB 9; Length 340000;  
Best Local Similarity 97.7%; Pred. No. 2.7e-80;  
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCCACTGCTGTGAGAAAGGTTCAATG 242  
DB 6374 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCCACTGCTGTGAGAAAGGTTCAATG 6433  
QY 243 TTGGTACAGTTCGGGGGCCCCCTGGAACGACAGCCTGCAAGAAACCGGTTTCTTC 302  
DB 6434 TTGGTACAGTTCGGGGGCCCCCTGGAACGACAGCCTGCAAGAAACCGGTTTCTTC 6493  
QY 303 CCAATAGGATGGCCCCCGGGGGGTCTCTTTCGACACCAAGATGGAGGAGTTCAGT 362  
DB 6494 CCAATAGGATGGCCCCCGGGGGGTCTCTTTCGACACCAAGATGGAGGAGTTCAGT 6553  
QY 363 CAGGCGAGATTTCAGGCTTGGCAGCATGGAGCAGGCGACAGACTGGGGAGTTCAGT 422  
DB 6554 CAGGCGAGATTTCAGGCTTGGCAGCATGGAGCAGGCGACAGACTGGGGAGTTCAGT 6613  
QY 423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGAAAGAGTGGCTCTCAGAGGGTCTG 482  
DB 6614 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGAAAGAGTGGCTCTCAGAGGGTCTG 6673  
QY 483 CACCCAGCCAGTCTGCAATGGGCGCTCTTGGCCTGTCGACAGAAATGAGAGCAGT 542  
DB 6674 CACCCAGCCAGTCTGCAATGGGCGCTCTTGGCCTGTCGACAGAAATGAGAGCAGT 6733  
QY 543 GCCGTGTGCGGAGCGGCGGAGCTCATCTGCTGACGCGCTGCGGCTTCCAC 602  
DB 6734 GCCGTGTGCGGAGCGGCGGAGCTCATCTGCTGACGCGCTGCGGCTTCCAC 6793  
QY 603 CTGGCTGCTGCTCCCTCCCTCCGTCGCGAGATCCCAAGTGGGACTGAGGCTTCCAC 662  
DB 6794 CTGGCTGCTGCTCCCTCCCTCCGTCGCGAGATCCCAAGTGGGACTGAGGCTTCCAC 6853  
RESULT 13  
AL954247 143192 bp DNA linear PRI 16-OCT-2003  
LOCUS Pan troglodytes chromosome 22 clone CH251-479113 map 22q22.3,  
DEFINITION complete sequence.  
ACCESSION AL954247  
VERSION AL954247.2 GI:37606037  
KEYWORDS HTG.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
1 (bases 1 to 143192)  
REFERENCE  
1 The Chimpanzee Chromosome 22 Sequencing Consortium  
DNA sequence of chimpanzee chromosome 22 and its evolutionary  
implications  
JOURNAL  
2 (bases 1 to 143192)  
Kube,M., Sudbrak,R., Mueller,I., Thiel,J., Klages,S., Borzym,K.,  
Heltman,K., Gimel,V., Beck,A., Ben Kahla,A., Lehrach,H.,  
Yaspo,M.L. and Reinhardt,R.  
TITLE Direct Submission  
JOURNAL Submitted (18-DEC-2002)  
COMMENT The Chimpanzee Chromosome 22 Sequencing Consortium consists of:  
\*Chinese National Human Genome Center at Shanghai,  
Shanghai, China;  
\*GBF, Dept. of Genome Analysis, Braunschweig, Germany;  
\*Institute of Molecular Biotechnology, Jena, Germany;  
\*KRIH Genome Research Center, Daejeon, Korea;  
\*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;  
\*National Institute of Genetics, Mishima, Japan;  
\*National Yang Ming University Genome Research Center, Taipei,

Taiwan;  
 \*RIKEN Genomic Sciences Center, Yokohama, Japan.  
 -----  
 Center: Max-Planck-Institute for Molecular Genetics  
 Center code: MPMG  
 -----  
 Project Information  
 Center clone name: CH251-479113  
 -----  
 Summary Statistics  
 Sequencing vector: pUC18; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 143190 bases at least Q40  
 Consensus quality: 143192 bases at least Q30  
 Consensus quality: 143192 bases at least Q20  
 Quality coverage: 25.4  
 -----  
 This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. -----  
 Neighboring sequence information:  
 This clone is overlapped by  
 PTB-051003  
 PTB-061A04  
 -----  
 SOURCE INFORMATION:  
 THE CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoji Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at  
 http://www.chori.org/bacpac/chimpanzee251.htm. The clone may be obtained from Pieter J. de Jong and coworkers  
 (http://www.chori.org/bacpac). VECTOR: pPARAC2.1  
 IMPORTANT: This sequence is not the entire insert of clone CH251-479113 It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 -----  
 FEATURES  
 source  
 1.143192  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /chromosome="22"  
 /map="22q22.3"  
 /clone="CH251-479113"  
 <1.14090  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-051003"  
 /note="overlapping clone"  
 143189..>143192  
 /organism="Pan troglodytes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9598"  
 /clone="PTB-061A04"  
 /note="overlapping clone"  
 -----  
 ORIGIN  
 Query Match 29.2%; Score 451.2; DB 9; Length 143192;  
 Best Local Similarity 96.2%; Pred. No. 4.7e-78;  
 Matches 462; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
 QY 183 AGGTGGAGAGACCTCTGGGCTTGAGCCCACTGCTGTGAGAGAGGTTTCATGTGG 242  
 DB 37910 AGGTGGAGAGACCTCTGGGCTTGAGCCCACTGCTGTGAGAGAGGTTTCATGTGG 37969

QY 243 TTGGTGTACAGTTTCGGGGGCCCCCTGGAGACGACCTCTGCAAGAAACCGGTTTCTTC 302  
 DB 37970 TTGGGCTTACAGTTTCGGGGGCCCCCTGGAGACGACCTCTGCAAGAAACCGGTTTCTTC 38029  
 QY 303 CCAATGAGGATGGCCCCCGGGGGGTGTCTGTGTGAGACCAAGATGGGAAACAGGTGGT 362  
 DB 38030 CCAATGAGGATGGCCCCCGGGGGGTGTCTGTGTGAGACCAAGATGGGAAACAGGTGGT 38089  
 QY 363 CAGGGCAGATTTTCAGGCCCTTGGACATGAGGACAGGACAGACTGGGAGTTCAAGT 422  
 DB 38090 CAGGGCAGATTTTCAGGCCCTTGGACATGAGGACAGGACAGACTGGGAGTTCAAGT 38149  
 QY 423 ACCCAGAGATGCTGCTGGGGGAGCTGTTTGGGAAAGAGGTGCTCTCAGAGGGTCTG 482  
 DB 38150 ACCCAGAGATGCTGCTGGGGGAGCTGTTTGGGAAAGAGGTGCTCTCAGAGGGTCTG 38209  
 QY 483 CACCCAGCCAGTGTGATGAGGCGCTCTGCTGTGCAGAGATGAGACGAGTGT 542  
 DB 38210 CACCCAGCCAGTGTGATGAGGCGCTCTGCTGTGCAGAGATGAGACGAGTGT 38269  
 QY 543 GCGGTGTGTGGGAGCGGCGGGAGCTCATCTGCTGTGACGCGCTGCTGGGCTTTCAC 602  
 DB 38270 GCGGTGTGTGGGAGCGGCGGGAGCTCATCTGCTGTGACGCGCTGCTGGGCTTTCAC 38329  
 QY 603 CTGGCTGCTCTTCCCTCTCCCTCGCGGAGATCCCAAGTGGGACCTGAGGTCTCCAGC 662  
 DB 38330 CTGGCTGCTCTTCCCTCTCCCTCGCGGAGATCCCAAGTGGGACCTGAGGTCTCCAGC 38389  
 -----  
 RESULT 14  
 AF128772 1656 bp mRNA linear ROD 26-OCT-2001  
 LOCUS AF128772  
 DEFINITION Mus musculus strain SJL autoimmune regulator (Aire) mRNA, partial cds.  
 ACCESSION AF128772  
 VERSION AF128772.1 GI:7108572  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 1656)  
 Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Deter,J.C. and She,J.X.  
 Chromosomal localization and complete genomic sequence of the murine autoimmune regulator gene (Aire)  
 Autoimmunity 31 (1), 47-53 (1999)  
 JOURNAL  
 MEDLINE  
 PUBMED 20059142  
 10593569  
 2 (bases 1 to 1656)  
 Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q., Deter,J.C., Davoodi-Serimiromi,A. and She,J.X.  
 Complete genomic sequence, gene structure and localization of the mouse Aire gene  
 JOURNAL  
 REFERENCES  
 3 (bases 1 to 1656)  
 Ruan,Q.G., Wang C.Y., Shi,J.D. and She,J.X.  
 Direct Submission  
 JOURNAL  
 Submitted (17-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road  
 FEATURES  
 source  
 1.1656  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="SJL"  
 /db\_xref="taxon:10090"  
 /chromosome="10"  
 /map="between D10Mit31 and D10Mit10"  
 1..>1656  
 /gene="Aire"  
 1..>1656  
 /gene="Aire"











Matches	580; Conservative	0; Mismatches	244; Indels	18; Gaps	3;
QY	452 TGGGAAGAGTGGCTCTCAGAGAGGTGTGCAACCCAGCCCAAGTCTGCATGGAGCTCTC				511
Db	858 TGAGCAGAAAGTGGGCGACAGGTGGGGTCTCCCTTCCATCCCTCCCAAGTGAAGC				917
QY	512 TTGGCTGTCCAGAAAGATGAGAGAGGTGTGCGGTGTGCGGGAAGGGGGAGTCAAT				571
Db	918 CCAAGTTAACAGAAAGACAGAGATGATGTGCGGTGTGCGGGAAGGGGGAGTCAAT				977
QY	572 CTGCTGTGACGGCTGTGCGCTCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTG				631
Db	978 CTGTTGTGACGGCTGTGCGCTCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTG				1037
QY	632 GATCCCAATGGGACCTGAGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				691
Db	1038 GATCCCAATGGGACCTTGTGAGATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				1097
QY	692 GCCCGGGGAG				751
Db	1098 GTCCAGCTGAGGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1157
QY	752 GCTTAAAGTGGCGGAG				811
Db	1158 ACTGAGGTGAGCTTCAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1217
QY	812 CACGACTCTTGTCTAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				871
Db	1218 TGTGTGTGTACATATGTAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT				1271
QY	872 GGAATCTCTGAGCTGTGCAACCCCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT				931
Db	1272 GGAAGCTTGAAGACTGT				1331
QY	932 TCTGT				991
Db	1332 ACCAAGCGCGCATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT				1391
QY	992 GCGCGGT				1051
Db	1392 TGT				1451
QY	1052 CTTGCGCTGCAATCTGTCTGAGAGAGCTG-----ACCCAGCCCTGTGTGTGTGTGT				1105
Db	1452 TCTCGCTGTCAATCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT				1511
QY	1106 GCTGTGCCCCAG				1159
Db	1512 TGTACCACTGT				1571
QY	1160 CAGTCAAG				1219
Db	1572 TGT				1631
QY	1220 CGATGAGATCTGTGAGTGGGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1279
Db	1632 TGAAGGAGATCTGTGAGTGGGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1691
QY	1280 CT 1281				
Db	1692 CT 1693				

RESULT 20

AF079536

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

1936 bp mRNA linear ROD 17-MAR-1999

Mus musculus autoimmune regulator (Aire) mRNA, complete cds.

AF079536.1 GI:4426598

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	1 (bases 1 to 1936)
AUTHORS	Wang,C.Y., Shi,J.D., Davoodi-Semirami,A. and She,J.X.
TITLE	Cloning of Aire, the mouse homologue of the autoimmune regulator (AIRE) gene responsible for autoimmune polyglandular syndrome type 1 (ASPI)
JOURNAL	Genomics 55 (3), 322-326 (1999)
MEDLINE	99168902
PUBMED	10049587
REFERENCE	2 (bases 1 to 1936)
AUTHORS	Wang,C.Y., Shi,J.D., Davoodi-Semirami,A. and She,J.X.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-1998) Pathology, Immunology and Laboratory Medicine, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES	Location/Qualifiers
source	1..1936
organism	"Mus musculus"
/mol_type	"mRNA"
/strain	"B6"
/db_xref	"taxon:10090"
/cissue	type="thymus"
1..1936	
/gene	"Aire"
58..1716	
/gene	"Aire"
/function	"transcription factor"
/codon_start	1
/product	"autoimmune regulator"
/protein_id	"AAD20444.1"
/db_xref	"GI:4426599"
/translation	"MAGDGMRLRLRLRHTETAVAIADAPFLHALADHVPBPKF DETLAKKRGKPOAFHALLSLTRDSGAILDPRLILPKYONLERYRRLHSILGFP KVDVLDNQRKGRKPLAGKRAVLPRLPTKRLALEPPATPRATLASVSRGSHLK TKPRKPDGNLSQHLPLGNGIQTMAVVRQAVTASGVPETRGAQVETPGDEKVG GSKKCIQVGEFPTPNKFEDEPSGMLKKARSSSLKDVPRAKQAVTIPGDEKVG QCGVPLPSLSEPOVQKNEDECAVCHDGDELICDGPAPFAHACISPLQIIPS GLMRSCCLQGVQONLSQPEYRSPPELPADPTPIVAGRSABKTRGSRBLKASSDA AVTYNLLAPHPALPLBLSALCPLLSANESRPGAPASRCSVCGDGEVLRCAHCA AAFHRRCHFPTRPAARPTNLRKSSASASTPPTPGTGAIVPTSGRPAFLAKVDDSD ASHDVLHRDLDSLNLHNSFDGILLQMAIQMSRPLAETPPPS"

## ORIGIN

Query Match 25.0%; Score 385.6; DB 10; Length 1936;

Best Local Similarity 68.9%; Pred. No. 5.7e-65;

Matches 580; Conservative 0; Mismatches 244; Indels 18; Gaps 3;

QY	452 TGGGAAGAGTGGCTCTCAGAGAGGTGTGCAACCCAGCCCAAGTCTGCATGGAGCTCTC				511
Db	870 TGAGCAGAAAGTGGGCGACAGCATGTGGGGTCTCCCTTCCATCCCTCCCAAGTGAAGC				929
QY	512 TTGGCTGTCCAGAAAGATGAGAGAGGTGTGCGGTGTGCGGGAAGGGGGAGTCAAT				571
Db	930 CCAAGTTAACAGAAAGACAGAGATGATGTGCGGTGTGCGGGAAGGGGGAGTCAAT				989
QY	572 CTGCTGTGACGGCTGTGCGCTCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				631
Db	990 CTGTTGTGACGGCTGTGCGCTCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				1049
QY	632 GATCCCAATGGGACCTGAGAGGTGTCTCAGCTGTCTGAGAGCAACAGTCTCAGAGGTGA				691
Db	1050 GATCCCAATGGGACCTTGTGAGATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG				1109
QY	692 GCCCGGGGAG				751
Db	1110 GTCCAGCTGAGGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1169
QY	752 GCTTAAAGTGGCGGAG				811
Db	1170 ACTGAGGTGAGCTTCAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				1229
QY	812 CACGACTCTTGTCTAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG				871



LOCUS AF128115 1933 bp mRNA linear ROD 29-FEB-2000  
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively  
spliced product 1b, complete cds.  
ACCESSION AF128115  
VERSION AF128115.1 GI:7108531  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1933)  
Ruan, Q., Wang, C.Y., Shi, J.D. and She, J.X.  
Expression and alternative splicing of the mouse autoimmune  
regulator gene (Aire)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1933)  
Ruan, Q., Wang, C.Y., Shi, J.D. and She, J.X.  
Direct Submision  
JOURNAL Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW  
Archer Road, Gainesville, FL 32610, USA  
FEATURES  
source  
1..1933  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="B6"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/map="between D10Mit31 and D10Mit10"  
1..1933  
/gene="Aire"  
58..1713  
/gene="Aire"  
/function="transcription factor"  
/note="alternatively spliced product 1b"  
/codon\_start=1  
/product="autoimmune regulator"  
/protein\_id="AF36460.1"  
/db\_xref="GI:7108532"  
/translation="MAGDQMLRRLRLRLRLTELAVIDANFLLHALADHDVPEDEK  
OETRLKEKGGCPQAFHLLSWLITPDSGLILDPWILPQVLEPSYSLHILDQFP  
KQVLDLNSKRKKPLAGPKAVLPPTPKRLKEPRATTPATLASKVSSEFSGHLK  
TPEPKPDNLSEOHPLANGIQTMAVQRAVTASGVPGTGAIVESILIQVYES  
GSKKCIQVGEFETPNKPEDPSGNKNAKSSSKPVRAQAQVTPGRDQKVG  
QCGVPLPLSEPVNONDEBACVHDGSELIICDGEPRFHLACLSPLOEIPSG  
LWRSCCLGVRQOQLSOPESVSPPLPATPPLVGLRSSEKTRSPSEIKASPA  
VTVYVNLAPHPAPLLEPSALCPILSAGNEGPGPAPSAKSCYCGCTGVLCAHAA  
AHRMCRHPTAARPGTINLRKSCSADSTPTPEGPBAYPTSGPRAPGLAKVGDUSA  
SHDPVLRDLSELINEHSFDGILQWAIQMSRPLAETPFSS"

ORIGIN  
Query Match 24.3%; Score 375.6; DB 10; Length 1933;  
Best Local Similarity 71.1%; Pred. No. Se-63;  
Matches 546; Conservative 0; Mismatches 204; Indels 18; Gaps 3;  
QY 526 AGAATGAGACGAGTGTGCGGTGTGCGGACGCGGAGAGTCACTGTGTGACGGCT 585  
Db 941 AGAACAGAGTAGTGTGCGGTGTGCGACGAGAGTAGTCACTGTGTGAGCGCT 1000  
QY 586 GCCCTGGGCGCTTCACCTGAGCTGCTGCTCCGCTCCGGAGATCCCAAGTGGGA 645  
Db 1001 GTCCCGGGGCTTCACCTGAGCTGCTGCTCCGCTCCGAGAGATCCCAAGTGGCC 1060  
QY 646 CCTGAGAGTCTCCAGTGTGCTGACAGCAACAGTCCAGAGAGTGAAGCCCGGAGAG 705  
Db 1061 TCTGAGAGTCTCTGCTGCTTCAGAGGAGAGTCCACAGAACTGTCCAGCTTAGAG 1120  
QY 706 AGCCCGGCGCCGAGAGCAACCCGCTGAGACCCCGGAGCTTAAAGTTCGAGCG 765  
Db 1121 TGTCCAGGCGCCCGGAGCTACTGCAAGAGACCCGATCTCTGAGGAGCTAGGCTT 1180  
QY 766 GAGAGAGAGTAAAGTCCACCTGGGGAACCCCTAGCCGAGATGACAGACTTGTCT 825  
Db 1181 CAGAGAAAACAGGGGCCCATCCAGGAGAGCTCAAGCGACAGCTGTGCTGTGACAT 1240

QY 826 ACAGACCTTGCAGGCTCCGCTCTGAGAGCCCGGTGCAGATCTGCTCGAGCC 885  
Db 1241 ATGTAACTGTGTCGCGCCGACCTTGACACTCTCTG-----CTGAGCTTACAGAC 1294  
QY 886 TGCACCCCTACTGTGTGTGTGAGTCTGAGAGTCAAGAACTGTGCTGTGAGCGCTT 945  
Db 1295 TGTGCTCTCTACTGAGTGTGTGGAATGAGAGGCGCGCAGAGTCCAGACCAAGCGCGAT 1354  
QY 946 GCGGGGTGTGCGAATGTATAGGACGTCTCGGTGTACTACTACGCGCGCTGCTTTC 1005  
Db 1355 GGAAGTGTGTGCGCAATGCAACGAGAGTGTTCGATGTGCAACATGTGCGCTTTC 1414  
QY 1006 ACTGAGCGTCCACTTCCAGCGGACCTCCGAGCCGAGAGGAGCCGTGCGCAGAT 1065  
Db 1415 ACTGAGCGTCCACTTCCAGCGGACCTCCGAGCCGAGAGGAGCCGTGCGCAGAT 1474  
QY 1066 CCGTCTCAGAGACGTG-----ACCCAGCCCTGTGTGAGGGGTGTGAGCCCGAGCC 1119  
Db 1475 CCGTCTCAGAGACGTGCTCCAGCGGACCTCCAGCGGAGAGCCAGCTGACCACTTTC 1534  
QY 1120 CCGCGCGCTGCGCCCTGTGAGCTGCGCA-----GATGACACTGCGACTACAGCCCG 1173  
Db 1535 GCGCCCGTCCAGACCTGTGAGCTTGTCCAGATGAGGAGACACTGTCTAGTACGACCTTG 1594  
QY 1174 CTTGACAGAGGATGACGTGAGTCCCTTCTGAGGAGACACCTTCATGAGCATCTTTC 1233  
Db 1595 TTCTCATATGAGGACACACTGAGATCTCTCTCTCATATGAGCATCTTTCAGCGCATCTTC 1654  
QY 1234 AGTGGCCATCCAGAGCATGAGCGCGTCCGCGAGCCCTTCCCTCT 1281  
Db 1655 AGTGGCCATCCAGAGCATGATCAGCGCGCTGCGAGACACCACTT 1702

RESULT 23  
HSLTPFK  
LOCUS HSLTPFK 2586 bp DNA linear PRI 29-APR-1996  
DEFINITION H.sapiens gene for liver type phosphofructokinase.  
ACCESSION X80853  
VERSION X80853.1 GI:1292864  
KEYWORDS phosphofructokinase.  
Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
Levanon, D., Brandeis, M., Bernstein, Y. and Groner, Y.  
Common promoter features in human and mouse liver type  
phosphofructokinase gene  
JOURNAL Biochem. Mol. Biol. Int. 35 (5), 929-936 (1995)  
MEDLINE 96018039  
PubMed 7549935  
REFERENCE 2 (bases 1 to 2586)  
Groner, Y.  
Direct Submision  
JOURNAL Submitted (04-AUG-1994) Y. Groner, The Weizmann Institute, Rehovot  
76100, ISRAEL  
FEATURES  
source  
1..2586  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone\_1fb="chromosome 21 enriched genomic library  
(cytometry 7;411-417;1986)"  
1..2586  
/product="phosphofructokinase"  
/note="liver type"  
447..616  
/note="1/2Aluv"  
/tpc\_family="Alu"  
1251..1597  
repeat\_region  
repeat\_region

misc\_binding /rpt\_family="Alu"  
2471..2526  
/bound\_molecy="Sp1"

## ORIGIN

Query Match 21.7%; Score 336; DB 9; Length 2586;  
Best Local Similarity 100.0%; Pred. No. 2.5e-55;

Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1210 AGCACACCTTCGATGAGCATCTCTGAGGAGCCATCCAGACATGAGCCCTCCGAGCC 1269  
135 AGCACACCTTCGATGAGCATCTCTGAGGAGCCATCCAGACATGAGCCCTCCGAGCC 194  
1270 CCTTCCCTCTCTGAGCCAGATGCGGAGCATGAGCTCTGATGAGAGATCTTGAA 1329  
195 CCTTCCCTCTCTGAGCCAGATGCGGAGCATGAGCTCTGATGAGAGATCTTGAA 254  
1330 GGACACCTCTCTGAGCTCTGAGGAGCCGAGGATCAAGAGGAGCAGGCGC 1389  
255 GGACACCTCTCTGAGCTCTGAGGAGCCGAGGATCAAGAGGAGCAGGCGC 314  
1390 ACCTCTGTCAGTCTGCGCTGAAACAGCTCTGTTCTGAGGAGCAGCAGCATCATG 1449  
315 ACCTCTGTCAGTCTGCGCTGAAACAGCTCTGTTCTGAGGAGCAGCAGCATCATG 374  
1450 TGCTTGGAATTAACCTGCGCCACCTTCTTACTGAGAGTCCCGGAGGCTCTCT 1509  
375 TGCTTGGAATTAACCTGCGCCACCTTCTTACTGAGAGTCCCGGAGGCTCTCT 434  
1510 TGCTTGGAATTAACCTGCGCCACCTTCTTACTGAGAGTCCCGGAGGCTCTCT 1545  
435 TGCTTGGAATTAACCTGCGCCACCTTCTTACTGAGAGTCCCGGAGGCTCTCT 470

RESULT 24 1744 bp mRNA linear ROD 29-FEB-2000  
AF128121  
LOCUS Mus musculus autoimmune regulator (Aire) mRNA, alternatively  
spliced product 2d, complete cds.

DEFINITION AF128121  
AF128121.1 GI:7108543

ACCESSION AF128121.1 GI:7108543

VERSION Mus musculus (house mouse)

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1744)  
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.  
Expression and alternative splicing of the mouse autoimmune  
regulator gene (Aire)

JOURNAL Unpublished  
2 (bases 1 to 1744)  
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.  
Direct Submision  
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW  
Archer Road, Gainesville, FL 32610, USA

FEATURES Location/Qualifiers  
1..1744  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/strain="B6"

gene /gene="Aire"

CDS /gene="Aire"

58..1524  
/gene="Aire"

1..1744  
/gene="Aire"

58..1524  
/gene="Aire"

1..1744  
/gene="Aire"

58..1524  
/gene="Aire"

1..1744  
/gene="Aire"

58..1524  
/gene="Aire"

1..1744  
/gene="Aire"

## ORIGIN

Query Match 11.3%; Score 174.6; DB 10; Length 1744;  
Best Local Similarity 69.9%; Pred. No. 7.1e-24;

Matches 269; Conservative 0; Mismatches 104; Indels 12; Gaps 2;

909 CCTGAGGTCAGAGAACCTGGCTCTGTGCGCGGTGGCGGAGTGGAGATGATG 968  
1129 CCTGAGAGAACCTGGCTCTGTGCGCGGTGGCGGAGTGGAGATGATGATG 1188  
969 GACGTGTCGAGTGACTGACCTGCGCGCTTTCACCTGAGGCTGACCTCCAGCC 1028  
1189 GAGGTGTCGAGTGACTGACCTGCGCGCTTTCACCTGAGGCTGACCTCCAGCC 1248  
1029 GGCACCTCCCGCGCGGAGCAGGCTGCGCTGACATCTGCTCAGAGACGTG----- 1082  
1249 GCGCGCGCGCGCGGAGCAGATCTCCGCTCAATCTGCTGACATCTGACTCC 1308  
1083 ACCCGAGCCCTGTGAGGAGGAGTGTGCGCCAGCCCGCGCGCTGAGGCTGAG 1142  
1309 AGCCAGGACACCGGAGGAGTGTGCGCCAGCCCGCGCGCTGAGGCTGAG 1368  
1143 GCCAA-----GATGACACTGCGACATGACGAGCCGCTCTGACAGGAGTGA 1196  
1369 GCCAAGTAGAGGAGCAGCTGCTGATGACGACCTGTTCTCATGAGGAGCAGCTGAG 1428  
1197 TCCCTTGAAGGAGCAGCTTTCATGAGGCTTTCATGAGGCTTTCATGAGGCTTTC 1256  
1429 TCCCTTGAAGGAGCAGCTTTCATGAGGCTTTCATGAGGCTTTCATGAGGCTTTC 1488  
1257 CGTCCGCGCGCGCGGAGCAGATCTCCGCTGACATCTGAGGCTGAGGCTGAG 1488  
1489 CGCGCGCGCGCGGAGCAGATCTCCGCTGACATCTGAGGCTGAGGCTGAG 1513

RESULT 25 1747 bp mRNA linear ROD 29-FEB-2000  
AF128120  
LOCUS Mus musculus autoimmune regulator (Aire) mRNA, alternatively  
spliced product 2c, complete cds.

DEFINITION AF128120  
AF128120.1 GI:7108541

ACCESSION AF128120.1 GI:7108541

VERSION Mus musculus (house mouse)

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1747)  
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.  
Expression and alternative splicing of the mouse autoimmune  
regulator gene (Aire)

JOURNAL Unpublished  
2 (bases 1 to 1747)  
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.  
Direct Submision  
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW  
Archer Road, Gainesville, FL 32610, USA

FEATURES Location/Qualifiers  
1..1747  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"

gene /gene="Aire"

58..1524  
/gene="Aire"

1..1747  
/gene="Aire"

58..1524  
/gene="Aire"

1..1747  
/gene="Aire"



VERSION	AF128118.1	GI:7108537
KEYWORDS		
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 1759)	
AUTHORS	Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.	
TITLE	Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 1759)	
AUTHORS	Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA	
FEATURES	Location/Qualifiers	
SOURCE	1..1759	

```

/mol_type="mRNA"
/strina="86"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1759
/gene="Aire"
58..1539
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 2a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AA036463.1"
/db_xref="GI:1708538"
/translation="MAGSGDMRLRLRLRLRTETAVIDSAFPLIPLADLADHYVEDKRE
OSTRLKREKCECPAPGALDLSMLLTDSGAILDFWRLTKPDNLERSYRLSLIDGFFP
TDPKDDNDSNRKRGKSLAPKPAVLPPTPRKRLAEPRATPLASKSVSPQSHL
TKPVKDDNLSQSLHLELGNGIOTMAASVORAVTVASGDPETRGAVGIIIOGFES
GGSKVGLGVGEFTLPKPKFEDNKKNAKSGSLKPVRAKGAQVITPRDEKVGAEI
GQGVVPLPLPSPPVONQNEDECAVCHDGEILLCDGCPRAFLACLSPLEIPEIS
GLMKSCCTCGRVQGNISQREVSPPPELPAETRGAPSAKSCVCGDGTETVLRCHACAA
AFNRHCFPTLAAAPGNTLKSCSDADSTPOTPEBAVPTSGRPAGLAKVGDLS
SHDVPLRRDDESLNLSHSPDGLIOWAIOQMSNPPLAETPPFS"

```

**ORIGIN**

Query Match	11.3%;	Score 174.6;	DB 10;	Length 1759;
Best Local Similarity	69.9%;	Pred. No. 7.1e-24;		
Matches 269; Conservative	0;	Mismatches 104;	Indels 12;	Gaps 2;

OY	909	CCTGAGGGTCAGCAAGAACTTGGATCTCTGGTGGCGCTTCGCGGGAGTGTGGAGAGATGTTACG	968
Db	1144	CTTCAGAGAGACCCCGGATTCAGACACAAAGCCGCGCATTCACATGTGTGTGGCGATGGCAAC	1203
OY	969	GACGTGCTCGCGGTGTACTACTGCGCGCGCTGCTTCACTGAGCGCTGCCACTTCCAGCC	1028
Db	1304	GAGGTGTTCGCGTGTGCACACTGTGCCCCCTGCTTCACTGCGCTGCACATTCCGCAAG	1263
OY	1029	GGCACTCTCCGGACCGGAGCGGGCTGTGGCTGCAGATCCTGTCTCAGAGACGTTG-----	1082
Db	1264	GCCGCGCGCGCGGCCGGGGAACCATTTCCGCTGCAAAATCTGTCTCTGCAGACTGCACCTGCC	1322
OY	1083	ACCCCAAGCCCCCTGTGAGGGGGGTGTGGCCCCCAAGCCCCCGCTGTGGCCCCCTGGAGCTT	1142
Db	1324	ACGCGAGGACACACCGGGGGAAGCTGTACCACTCTGTGGGCCCGCTTCAGACACTTGGAGCTT	1383
OY	1143	GCCAA-----GGATGACACTGCGCAGTACAGAGCCCGCTTGTGCACAGGAGATGACTCTGGAG	1196
Db	1384	GCCAAAGTATGGGAGACGACTCTGTCTAAGTACACACCTGTTCTACATAGGGAAGACCTGTGAG	1443
OY	1197	TTCCTTCTGAGGAGACACCTTTCGATGGCATCTGCAGTGGGCACTCAAGACATGGCC	1256
Db	1444	TCCCTCTTCATATGACACTTTTGAACGCAATCTGCAGTGGGCACTCAAGACATGTGCA	1503

Qy	1257	CGTCCGGCGGCCCCCTTCCCTCT	1281
Db	1504	CGCCCGTGGCGGAGACACACCT	1528

RESULT	28
LOCUS	AF128124
DEFINITION	AF128124 1625 bp mRNA linear ROD_29-FEB-2000 Mus musculus autotimmune regulator (Aire) mRNA, alternatively spliced product 3c, complete cds.

ACCESSION	AF128124
VERSION	AF128124.1
	GI:7108549

KEYWORDS	SOURCE
Mus musculus (house mouse)	.

ORGANISM

REFERENCE  
1 (bases 1 to 1625)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus

AUTHORS	Ruan, Q. G., Wang, C. Y., Shi, J. D. and She, J. X.
TITLE	Expression and alternative splicing of the mouse autoimmune

# 101

JOURNAL	unpublished
REFERENCE	2 (bases 1 to 1625)
AUTHORS	Ruan, Q. G., Wang, C. Y., Shi, J. D. and She, J. X.

**TITLE** Direct Submission  
**SUBMITTED** (12-FEB-1999) Pathology, University of Florida, 1600 SW  
**JOURNAL** Archer Road, Gainesville, FL 32610, USA

## FEATURES

### source

```

/organism="Mus musculus"
/mol_type="mrna"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/mpg="between D10Mit31 and D10Mit10"
1..1625
gene

```

**CDS**

```

/note="alternatively spliced product 3c"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAJ36469.1"
/db_xref="GI:7108550"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLLHALADHVVDEKRE
OELLRLKEKGCPOAFALLSMILLTRSGAILDFWRLLFDYNERLSRLSLIDGPE
KVDVLDNOSRKGRKPLAGPKAAVLPPTPTKRALEBRATPPATLAKSVSSPSHLK
TKPKPKDGNLIESQHLPLNGDIQTMAVSRAVTVAGDVGTGAVEGLIIQQVFSS
GRSKKCIQVGGEFTPTPKFEDPSGNLGNKRSAGSLVAVRAKQAQGRDEKVCQQCC
VPLPPLSRBPVONKNEDCAVCHDGGELICDGGCRRAHFIACLSPLDIEPSGLMWR
SCCCLQSRVQONLSQPEVSRPELPETPPQSPFLQILLCRLDSHARTGTSCTHLMAR
SSTWACGGRGLC"

```

**ORIGIN**

Query Match	10.3%	Score 159;	DB 10;	Length 1625;
Best Local Similarity	70.3%	Pred. No. 7.8e-21;		
Matches 213; Conservative	0;	Mismatches 90;	Indels 0;	Gaps 0;

452 TGGGAGGAGGTGGCTTCAGGAGGTGCTGCACCCAGGCCAGTCTGCATGGGCGTCTC 511

Db 858 TGAGCAGAAAGTGGGCACGACAGTGTGGGGTTCCTCCCTTCATCCCTCCACAGTGAGCC 911

512 TTGCTGTGTCAGAGAATGAGGACGAGTGTGCGGTGTGTCCGGACGGCGGGGAGCTCAT 572

db 918 CCAAGTTAACCAAGAAAGAGATGATGTCCTGTCACGACGAGGTGACTCAT 977

572 CTGCTGTGACGGCTGCCCCCTGGGGCCCTTCCACCGGGCCCTGCTGTCCCTCCGCTCCGGGGA 631

db 978 CTGTTGTACGGCTGTCCCGGGCCTTCCACCTGGCTTGCTGTGCCACCTCTGCAGGA 1033

632 GATCCCAAGTGGGACCTGGAGGTGCTCCAGCTGCCTGCAGGCAACAGTCCAGGAGGTCCA 631



Db	1038	GATCCCCAGTGGCTCTTGGAAATGCTCTGCTGCTCCAGGGGCGAGTGCCAAAGAACT	1097
QY	692	GCCCCGGGACAGAGAGCCCGGCGCCCGACAGAGCCACCGGTGAGAGACCCCGCTCCCCGGG	751
Db	1098	GTCACAGCTGAGGTGTCCAGGCCCCCGAGACTACTGACAGAGACCCCGGACCAATCTCC	1157
QY	752	GCT 754	
Db	1158	GCT 1160	
RESULT 29			
LOCUS	AF128122	1637 bp	mRNA linear ROD 29-FEB-2000
DEFINITION	Mus musculus autoimmune regulator (Aire) mRNA, alternatively spliced product 3a, complete cds.		
ACCESSION	AF128122		
VERSION	AF128122.1	GI:7108545	
KEYWORDS			
SOURCE			
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1637)		
TITLE	Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.		
JOURNAL	Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1637)		
TITLE	Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.		
JOURNAL	Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA		
FEATURES			
source	Location/Qualifiers		
	1..1637		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="B6"		
	/db_xref="taxon:10090"		
	/chromosome="10"		
	/map="between D10Mit31 and D10Mit10"		
	1..1637		
	/gene="Aire"		
	58..1287		
	/gene="Aire"		
	/function="transcription factor"		
	/note="alternatively spliced product 3a"		
	/codon_start=1		
	/product="autoimmune regulator"		
	/protein_id="AAP36467.1"		
	/db_xref="GI:7108546"		
	/translation="MAGSDGMLRRLRLRLRRLREIVAVDAISAPFLILHALDDHVDKDFQETRLRKEKCECPQAFHALLSWMLTRBSGALIDPWRLIFDYNLERISRLHSITIDPRLKDVLNDSRKRRKPLIAGPKAVLPPRPPTKRLSEBPATPPTATLVSSVSSPESHUKTKPPEKPDGNLSQHLPLNGNIQTMAASVQRAVTVASDVGATGAVAGGLIDQFESSGRSKCIQVGEFETYPNFKEDPDSGLNKKASGSLKPVVAKAGAQVTIPGRDOKVGGCCVPLPLSPSEPVQNKDEBCAVCHDGEGLICDGCPRAFHACLSPLOEISSGLMGSCTCIGRVQOONLSQPEVSRPELPATPPQSPPLQILGLDHAHNTGNSCHLMAISSTMACCGRRRLC"		
gene			
CDS			
ORIGIN			
Query Match	10.3%;	Score 159;	DB 10; Length 1637;
Best Local Similarity	70.3%;	Pred. No. 7.8e-21;	
Matches 213; Conservative	0;	Mismatches 90;	Indels 0; Gaps 0;
QY	452	TGGGAGAGAGGTGTGCTTCAGGAGGCTCTGCACCCACGCGCCAGTCTGCATGGCGCTTC	511
Db	870	TGAGCGAGAAAGTGGGCGACAGTGTGGGGTTCCTCCCTTCATCTCCCTCCCACTGAGCC	929
QY	512	TTGCGTGTGGCCAGAAAGATGAGAGAGAGTGTGCGGTGTGTGCGGAGCGCGGAGAGTCAT	571
Db	930	CCAGGTAAACAGAGAAACAGAGATGAATGTGTGCGGTGTGCGGACAGCGAGGTGAGCTCAT	989

QY	572	CTGCTGTGACGGCGTCCCTCGAGGCGTTTCACACTTGAGCTGCTGCCTCCCTCCGCTCCGGGA	631
Db	990	CTGTTGTGACCGGCTGTCCCCGGGCGTTTCACACTTGAGCTGCTGCCTCCCACTCTGCAGAGA	1049
QY	632	GATCCCCGAGTGGAGACTGAGAGTGTCTCCAGCTGTGCTTCAGAGCAACAGTCCAGAGAGTGA	691
Db	1050	GATCCCCGAGTGGCGCTCTGTGAGATGTCTCTGTGTGCTTCAGAGGCGAGATTCACAAGAACT	1109
QY	692	GCCCCGGGACAGAGAGGCCCGCGGCCCCAGAGACCACCGTGGAGACCCCGCTCCCGCCGGG	751
Db	1110	GTCCACAGCCTGAGGTGTCTCAGAGCCCCCGAGAGTACTGACAGAGACCCCGAACCATCTCC	1169
QY	752	GCT 754	
Db	1170	GCT 1172	
RESULT 30			
AF128125			
LOCUS	AF128125	1622 bp	mRNA linear ROD 29-FEB-2000
DEFINITION	Mus musculus autoimmune regulator (Aire)	mRNA, alternatively spliced product 3d, complete cds.	
VERSION	AF128125		
KEYWORDS	AF128125.1 GI:7108551		
ORGANISM	Mus musculus (house mouse)		
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1622)		
AUTHORS	Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.		
TITLE	Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1622)		
AUTHORS	Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA		
FEATURES	Location/Qualifiers		
source	1..1622		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="B6"		
	/db_xref="taxon:10090"		
	/chromosome="10"		
	/map="between D10Mit31 and D10Mit10"		
	1..1622		
	/gene="Aire"		
	58..1272		
	/gene="Aire"		
	/function="transcription factor"		
	/note="alternatively spliced product 3d"		
	/codon_start=1		
	/product="autoimmune regulator"		
	/protein_id="AAF36470.1"		
	/db_xref="GI:7108552"		
	/translation="MAGDGLRLRLHRTETLAVIDSAFPLHLADHDVPEDEKFEOTLRKEKGCCPOAFHALSWLITRSGALILDPWRLLPFDVNLERSRLSHLDGPPKDVLNDSRKRRKFLIAPKAAYVPRPTPKALEBRATPPATLAKSYSSRSNSHKTKDKPRPDGNLSQHLPVGNGIQTMASVGPRAVTAVAGDVGRGAVEGILLQVFESGRSKCIQVGEFFTPNKFPDPSGNLNKRKSGLSLPVVRAGAQRDRQKVGOQCVPPLSPISPEFOVNEDECVAGCDGELLCCGCPAFHLACISPLOEIPGLWRCSCICGQVOONLSQPVSRRPELPALRPDQSPIDQLIRLSDSHARHTRGSTLMAPSSTMACGSRGLC"		
CDS			
gene			
ORIGIN			
Query Match	9.6%	Score 149;	DB 10; Length 1622;
Best Local Similarity	78.2%;	Pred. No. 6.9e-19;	
Matches	179;	Conservative 0;	Mismatches 50; Indels 0; Gaps 0;
526	AGAAATGAGACGAGTGTGCCGTGTGTGGGAGCGGCGGAGACTCATCTGCTGTGACGGCT	585	

QY	526	AGAA	ATGAGG	ACGAGT	GCCTGT	ATCGG	ACGCGGGG	AGCTAT	CTCGTGT	ACGCT	585
Db	941	AGAA	CGAGG	ATGAGT	GCCTGT	CCACCA	CGAGAGT	AGCTAT	CTGTGT	ACGCT	1000

### FEATURES

#### Location/Qualifiers

```

source
1. .9876
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match      6 6%; Score 101.8; DB 6; Length 9876;
Best Local Similarity 69.2%; Pred. No. 8e-10;
Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 AGAAGAGTGGAGTCTTCTCAGGCTCTTAAGAGCATGGGCTGTGCTGACGCG 60
Db 491 AAAAAAAAAATAATCTTCTCAAACTTTAAAAACATAGCTTTAATCCAACTATACCG 432
QY 61 CTGCTCTGAGTGGGCGCGGCGGCGCGCGCGCTGCTATAGCCAGAGGTCAAG 120
Db 431 CTAAGCTCAACTAAACCCGTAATAATAAACGAAAGCGCCCTACTATACCAAAAAATCAGAA 372
QY 121 ATCCACTGGGAGTGCATCTTGTGCTCCGAGCATGCTTCTTAATGGGGTAGAA 180
Db 371 ATCCACTAAAAAATACCATCTACTTGTGCTCCCAACATTAATTTCTAATAATAATAA 312
QY 181 GCAGGTGCGGAGAGACCTCCC 201
Db 311 ACAATATAAAAAATACTTACC 291

RESULT 34
BV089377      581 bp      DNA      linear      STS 15-OCT-2003
LOCUS      RPAMMS00001273 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV089377
VERSION      BV089377.1 GI:37666856
KEYWORDS      STS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 581)
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
Mus musculus SNPs
Unpublished (2003)

TITLE      JOURNAL
COMMENT      Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted
Primer B: No primer submitted
Location/Qualifiers
1. .581
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="10-11206-10631-CAA01111727.1.1.11206"
/clone_lib="Roche Palo Alto"
/notes="SNPs developed from assay sequences derived from 15
different strains-of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D-H2/obnJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J,
NBL/MPJ, NZB/B1mJ, NZW/Lac, SPRET/Ei..."
<1. .>581

STS
ORIGIN
Query Match      6.3%; Score 96.6; DB 11; Length 581;
Best Local Similarity 85.6%; Pred. No. 1.3e-08;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

```

```

QY 502 TGGAGGCTCTTACCGTGGCCAGAGAAATGAGACAGAGTGCCTCGTGTCCGACGCG 561
Db 274 TGGAGTCCCTCTTGCTTG-GTCAAGAGAGAGATGATGTCCGTGTGCCACGACGAG 332
QY 562 GGAAGCTCATCTGCTGTGACGCGCTCCCTCGGCGCTTCCACCTGCGCTGCTCCCTC 621
Db 333 GTGAGCTCATCTGTTGTGACGCGCTGCCCCGGCGCTTCCACCTGCGCTGCTCCCGAC 392
QY 622 CGCTCCGGAGATCCCCAG 640
Db 393 CTCTGACGAGATCCCCAG 411

RESULT 35
AF105002      18351 bp      DNA      linear      ROD 26-OCT-2001
LOCUS      Mus musculus autoimmune regulator (Aire) gene, complete cds.
DEFINITION      AF105002
ACCESSION      AF105002
VERSION      AF105002.1 GI:5669675
KEYWORDS      .
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 18351)
Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.Q., Huang,Y.Q.,
Dettler,J.C. and She,J.X.
Chromosomal localization and complete genomic sequence of the
murine autoimmune regulator gene (Aire)
Autoimmunity 31 (1), 47-53 (1999)
20059142
MEDLINE      10593569
JOURNAL      2 (bases 1 to 18351)
Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.Q., Huang,Y.Q.,
Dettler,J.C. and She,J.X.
Direct Submission
Submitted (06-NOV-1998) Department of Pathology, Immunology and
Laboratory Medicine, University of Florida, 1600 SW Archer Rd.,
Room D6-15, Gainesville, FL 32610, USA
Location/Qualifiers
1. .18351
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10"
<4918. .>18138
/gene="Aire"
/join("<4918. .5052,5365. .5539,5769. .5925,6219. .6292,
6740. .6853,8274. .8422,8611. .8691,10373. .10488,
11577. .11676,12530. .12706,13659. .13780,13879. .13987,
16489. .16557,18058. .>18138)
/gene="Aire"
/product="autoimmune regulator"
/join(4918. .5052,5365. .5539,5769. .5925,6219. .6292,
6740. .6853,8274. .8422,8611. .8691,10373. .10488,
11577. .11676,12530. .12706,13659. .13780,13879. .13987,
16489. .16557,18058. .18138)
/gene="Aire"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAD46421.1"
/db_xref="GI:5669676"
/translation="MAQGDMMRLRLRLHRTETAVADISAPRLALADHDVDPEDKF
DELTLKEBGPQAFHLLSLWLTTRDGAIIIDPRITLKYDNLERYSLRISIDGFP
KVDVNDKQRKGRPLAGKAAVLPRLPKRALEPRATPATLASSVSPSSHLK
TPRPKDPGNLESQHLPLGNGIQTAAVORAVTVASGDPVPTAGVATLILQOVES
GSKKCIQVGFYTPNKFEDPSGLKKKASGSLAKEVPAKGAQVITPGRDEQVGS
QOCQGPLPSLSEPPVONKNEDECAVCHDGEELICDGPAPFLACTLSPLOEIPS
GLMRSCTCQGRVQNTLSQPRYSRPELPATRPITVIGURASAEKTRGSRRLKASDA
AVTVNLAPHPAPLEPSALCPILSKGNBRPAPASRSCVCGDGTETVLRCHKA
AAFWHRCHFTPAARPTNLRCKSCSADSTPTPTGPAVPTSGRPAVGLAKVGDG
ASHDVLVLRHDLTSLNHSFDGILQWAIQSMRSLAETPPSS"
CDS
ORIGIN

```



```

exon                /number=6
                   /gene="Aire"
                   /number=7
intron              6832..8505
                   /gene="Aire"
                   /number=7
repeat_unit         7002..7159
                   /gene="Aire"
                   /rpt_family="SINE/B4"
repeat_unit         7027..7160
                   /gene="Aire"
                   /rpt_family="SINE/Alu"
repeat_unit         complement(7758..7893)
                   /gene="Aire"
exon                /rpt_family="SINE/B2"
                   8506..8621
                   /gene="Aire"
intron              /number=8
                   8622..9709
                   /gene="Aire"
repeat_unit         9312..9458
                   /gene="Aire"
                   /rpt_family="SINE/Alu"
exon                9710..9809
                   /gene="Aire"
intron              /number=9
                   9810..10660
                   /gene="Aire"
repeat_unit         complement(10146..10356)
                   /gene="Aire"
repeat_unit         /rpt_family="SINE/B2"
                   complement(10359..10551)
                   /gene="Aire"
exon                /rpt_family="SINE/B2"
                   10661..10837
                   /gene="Aire"
intron              /number=10
                   10838..11786
                   /gene="Aire"
repeat_unit         /number=10
                   11060..11182
                   /gene="Aire"
repeat_unit         /rpt_family="SINE/B4"
                   complement(11453..11562)
                   /gene="Aire"
exon                /rpt_family="LINE/L1"
                   11787..11908
                   /gene="Aire"
intron              /number=11
                   11909..12004
                   /gene="Aire"
exon                /number=11
                   12005..12113
                   /gene="Aire"
intron              /number=12
                   12114..14604
                   /gene="Aire"
repeat_unit         complement(12262..12378)
                   /gene="Aire"
repeat_unit         /rpt_family="SINE/Alu"
                   complement(12277..12408)
                   /gene="Aire"
repeat_unit         /rpt_family="LTR/MaLR"
                   complement(12415..12431)
                   /gene="Aire"
repeat_unit         /rpt_family="SINE/B2"
                   complement(12432..12626)
                   /gene="Aire"
                   /rpt_family="SINE/B2"

```

```

repeat_unit         complement(12627..12769)
Query Match        6.3%; Score 96.6; DB 10; Length 18616;
Best Local Similarity 85.6%; Pred. No. 7.4e-09;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;
QY                502 TGAGGCTCTTGGCTGTCACAGAAATGAGAGACGAGTGTCCGTGTGTCCGACGGCG 561
                   |||||
DB                8484 TGGGTCCCTCTTGCCCTG-GTCAGAGAGACGAGATGAGTGTCCGTGTGCCACGACGAG 8542
QY                562 GCGAGCTATCTGCTGTGTGACGGCTGCTCCGCGGCTTGCACCTGCGCTGTGCTCCCTC 621
                   |||||
DB                8543 GTGAGCTATCTGTGTGTGACGGCTGTCCCGGGCTTCCTGCACCTGTGCTGCCAC 8602
QY                622 CGCTCGGAGATCCCGCAG 640
                   |||||
DB                8603 CTCTGAGAGATCCCGCAG 8621
RESULT 37
AF073797
LOCUS              AF073797 46872 bp DNA linear ROD 05-AUG-1999
DEFINITION        Mus musculus cosmid MPMGc12112287 containing the syntenic region
ACCESSION         AF073797
VERSION           AF073797.1 GI:4091972
KEYWORDS          HTG.
SOURCE            Mus musculus (house mouse)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE         1 (bases 1 to 46872)
AUTHORS           Aaltonen,U., Björnsen,P., Perheentupa,J., Horelli-Kuitman,N.,
                  Palotie,A., Peltonen,L., Lee,Y.S., Francis,F., Hennig,S., Thiel,C.,
                  Lehnach,H. and Yaspo,M.-L.
                  An autoimmune disease, APECED, caused by mutations in a novel gene
                  featuring two PHD-type zinc-finger domains. The Finnish-German
                  APECED Consortium. Autoimmune
                  Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy
                  Nat. Genet. 17 (4), 399-403 (1997)
JOURNAL           98061087
MEDLINE           9398840
PUBMED            9398840
TITLE             2 (bases 1 to 46872)
AUTHORS           Blechschmidt,K., Schweizer,M., Wertz,K., Poulsen,R.,
                  Christensen,H.M., Rosenthal,A., Lehnach,H. and Yaspo,M.L.
                  The mouse Aire gene: comparative genomic sequencing, gene
                  organization, and expression
                  Genome Res. 9 (2), 158-166 (1999)
JOURNAL           99148139
MEDLINE           10022980
PUBMED            10022980
TITLE             3 (bases 1 to 46872)
AUTHORS           Blechschmidt,K., Huang-WY,C., Nordsiek,G., Drescher,B.,
                  Rosenthal,A. and Yaspo,M.-L.
                  Direct Submission
                  Submitted (22-JUN-1998) Genome Analysis, Institute of Molecular
                  Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES          location/Qualifiers
source            1..46872
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /strain="129/Ola"
                  /db_xref="taxon:10090"
                  /note="Cosmid MPMGc12112287 originating from cosmid
                  library number 121 (RZPD-Berlin). Cloning vector Lawrlist
                  7."
exon              complement(156..327)
                  /note="Genscan, score = 14.95%, comment = Internal_exon
                  172 bp frame: 2 phase: 1"
exon              /evidence=not experimental
                  complement(156..327)
                  /note="GRABL, score = 99.000%, comment = excellent"
                  /evidence=not experimental
                  complement(156..316)
exon

```



exon 13336, 13416  
/gene="Mouse homologue of the human APECD gene (AIRE);

Query Match 6.3%; Score 96.6; DB 10; Length 46872;  
Best Local Similarity 85.6%; Pred. No. 6.4e-09;  
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

QY 502 TGGGGCTCTTGTGCTGTCGCGAAGATGAGACGAGTGCCTGCTGTCGGACGGCG 561  
DB 15069 TGGGTCCTCTTGCTGCTG-GTCAGAGAACGAGATGATGTCCTGTCACGACGAG 15127

QY 562 GGGAGCTATCTGCTGTGACGCGCTGCGGCTTCACCTGACCTGCTGCTGCTC 621  
DB 15128 GGAAGCTATCTGCTGTGACGCGCTGCTGCTGCTTCACCTGCTGCTGCTCCAC 15187

QY 622 CGCTCCGGAGATCCCGCAG 640  
DB 15188 CTCTCAGAGATCCCGCAG 15206

RESULT 38  
AC015891 158049 bp DNA linear HTG 20-NOV-2002  
LOCUS Mus musculus chromosome 10 clone RP21-522L13 map 10, \*\*\* SEQUENCING  
DEFINITION IN PROGRESS \*\*\*, 6 unordered pieces.  
ACCESSION AC015891.16 GI:25140119  
VERSION AC015891  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEPIN.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 158049)  
Birren, B., Nussbaum, C. and Lander, E.  
Mus musculus chromosome 10, clone RP21-522L13  
2 (bases 1 to 158049)  
2 (bases 1 to 158049)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barne, N., Beckert, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Liu, C., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McQuirk, A., McKernan, K., McLaughlin, D., Melgrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 158049)  
Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barne, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chararo, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lander, E., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menue, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riese, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (20-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Nov 20, 2002 this sequence version replaced gi:25046447. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L761  
Center clone name: 522\_L\_13  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 58675: contig of 58675 bp in length  
\* 58676 58775: gap of 100 bp  
\* 58776 75633: contig of 16858 bp in length  
\* 75634 75733: gap of 100 bp  
\* 75734 78147: contig of 2414 bp in length  
\* 78148 78247: gap of 100 bp  
\* 78248 107125: contig of 28878 bp in length  
\* 107126 107225: gap of 100 bp  
\* 107226 143742: contig of 36517 bp in length  
\* 143743 143842: gap of 100 bp  
\* 143843 158049: contig of 14207 bp in length.  
Location/Qualifiers  
1. 158049  
/organism="Mus musculus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/map="10"  
/clone="RP21-522L13"  
/clone\_1fb="RP21-522L13"  
/clone\_1fb="RP21-522L13" Female Mouse PAC"  
ORIGIN  
Query Match 6.3%; Score 96.6; DB 2; Length 158049;  
Best Local Similarity 85.6%; Pred. No. 5.2e-09;  
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;  
QY 502 TGGGGCTCTTGTGCTGTCGCGAAGATGAGACGAGTGCCTGCTGTCGGACGGCG 561  
DB 152857 TGGGTCCTCTTGCTGCTG-GTCAGAGAACGAGATGATGTCCTGTCACGACGAG 152915

QY 562 GGGAGCTATCTGCTGTGACGCGCTGCGGCTTCACCTGACCTGCTGCTGCTC 621  
DB 152916 GGAAGCTATCTGCTGTGACGCGCTGCTGCTGCTTCACCTGCTGCTGCTCCAC 152975

QY 622 CGCTCCGGAGATCCCGCAG 640  
DB 152976 CTCTCAGAGATCCCGCAG 152994

RESULT 39  
AC138672/c 190019 bp DNA linear HTG 22-MAR-2003  
LOCUS Mus musculus chromosome 10 clone RP21-411J14 map 10, WORKING DRAFT  
DEFINITION SEQUENCE, 13 unordered pieces.  
ACCESSION AC138672.3 GI:29150492  
VERSION AC138672  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 190019)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 10, clone RP23-411J14  
 Unpublished  
 2 (bases 1 to 190019)  
 Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barne, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fairo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (15-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 190019)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barne, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Cornu, B., Dearrellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fairo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 22, 2003 this sequence version replaced g1:28394987.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 411\_J14  
 Center clone name: L28928  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 186901 bases at least Q40  
 Consensus quality: 187957 bases at least Q30  
 Consensus quality: 188318 bases at least Q20  
 Insert size: 188819; sum-of-coverage  
 Quality coverage: 7.6 in Q20 bases; sum-of-coverage

-----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 1 11865: contig of 11865 bp in length  
 \* 11865: gap of 100 bp  
 \* 11965: contig of 3106 bp in length  
 \* 15071: gap of 100 bp  
 \* 15072: contig of 2227 bp in length  
 \* 15172: gap of 100 bp  
 \* 17399: contig of 1594 bp in length  
 \* 17992: gap of 100 bp  
 \* 19093: contig of 3644 bp in length  
 \* 19193: gap of 100 bp  
 \* 22837: gap of 100 bp  
 \* 22937: contig of 8173 bp in length  
 \* 31109: gap of 100 bp  
 \* 31210: contig of 14746 bp in length  
 \* 45955: gap of 100 bp  
 \* 45956: gap of 100 bp  
 \* 46056: contig of 24415 bp in length  
 \* 70470: gap of 100 bp  
 \* 70471: contig of 26331 bp in length  
 \* 70571: gap of 100 bp  
 \* 97501: contig of 20264 bp in length  
 \* 97502: gap of 100 bp  
 \* 97502: contig of 20264 bp in length  
 \* 117865: gap of 100 bp  
 \* 117866: contig of 22703 bp in length  
 \* 117965: gap of 100 bp  
 \* 140669: gap of 100 bp  
 \* 140768: contig of 26336 bp in length  
 \* 167069: gap of 100 bp  
 \* 167105: contig of 22815 bp in length  
 \* 167205: Location/Qualifiers  
 1. 190019  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="10"  
 /map="10"  
 /clone="RP23-411J14"  
 /clone\_lib="RP23-411J14"  
 1. 11865  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 11965..15071  
 /note="assembly\_fragment"  
 15172..17398  
 /note="assembly\_fragment"  
 17499..19092  
 /note="assembly\_fragment"  
 19193..22836  
 /note="assembly\_fragment"  
 22937..31109  
 /note="assembly\_fragment"  
 31210..45955  
 /note="assembly\_fragment"  
 46056..70470  
 /note="assembly\_fragment"  
 70571..97501  
 /note="assembly\_fragment"  
 97602..117865  
 /note="assembly\_fragment"  
 117966..140668  
 /note="assembly\_fragment"  
 140769..167104  
 /note="assembly\_fragment"  
 167205..190019  
 /note="assembly\_fragment"  
 clone\_end:77  
 vector\_side:right

```

ORIGIN
Query Match      6.3%; Score 96.6; DB 2; Length 190019;
Best Local Similarity 85.6%; Pred. No. 5,1e-09;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

OY 502 TGGGGGCTCTTGGCCCTGCGCAGAGATGAGACGAGTGCCTGCTGTCGGGAGGGG 561
    |||||
DB 59846 TGGGTCTCTTGTGCTCTG-GTGAGAACAGAGATAGTGTGCGGTGTCCACGACGAG 59788
    |||||

OY 562 GGGAGCTCATCTGCTGTGAGCGGTGCGCTTCGAGCCTGAGCTGCTGCTCTC 621
    |||||
DB 59787 GTGAGCTCATCTGTTGTGACGGGTGTCCCGGGCCTTCACCTGCTGCTGCTCCAC 59728
    |||||

OY 622 CGCTCGGAGATCCCGAG 640
    |||||
DB 59727 CTCTGACGAGATCCCGAG 59709
    |||||

RESULT 40
AC108592/c 217688 bp DNA 1linear HTG 15-NOV-2002
LOCUS
DEFINITION
AC108592
AC108592.5 GI:25006698
HTG, HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 217688)
Muzny,D,Marie, Metzker,M, Lee, A, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsdorfs, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Balwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devalla, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garsa, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R. J., Lu, X., Ma, T.,
Maheshwari, M., Maindarche, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawlin, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackeme, O., Okunolu, G., Olarinmashoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, R., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, R., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. U.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Tjorja, Z., Umant, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 217688)
Worley, K. C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217688)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23270241.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPMS
Center clone name: CH230-144C19
----- Summary Statistics
Assembly program: PHRAP; version 0.990329
Consensus quality: 177651 bases at least Q40
Consensus quality: 182606 bases at least Q30
Consensus quality: 185802 bases at least Q20
Estimated insert size: 178707; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Setmaped insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 161680: contig of 161680 bp in length
* 161780: gap of unknown length
* 161781: contig of 9109 bp in length
* 170889: gap of unknown length
* 170890: gap of unknown length
* 170990: contig of 6800 bp in length
* 177789: gap of unknown length
* 177790: contig of 11792 bp in length
* 177889: gap of unknown length
* 177890: contig of 11792 bp in length
* 189681: gap of unknown length
* 189781: gap of unknown length
* 189782: contig of 21519 bp in length
* 211301: gap of unknown length
* 212789: contig of 1389 bp in length

```

```

* 212790 212889: gap of unknown length
* 212890 214330: contig of 1441 bp in length
* 214331 214430: gap of unknown length
* 214431 215739: contig of 1309 bp in length
* 215740 215839: gap of unknown length
* 215840 217688: contig of 1849 bp in length.
Location/Qualifiers
Source
1. 217688
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-144C19"
7490..10042
misc_feature
/note="wgs_contig"
10093..11330
misc_feature
/note="wgs_contig"
71480..74410
misc_feature
/note="wgs_contig"
87246..89004
misc_feature
/note="wgs_contig"
15433..155684
misc_feature
/note="wgs_contig"
160679..161680
misc_feature
/note="wgs_contig"
170990..172417
misc_feature
/note="wgs_contig"
175878..177789
misc_feature
/note="wgs_contig"
177890..179119
misc_feature
/note="wgs_contig"
189782..190939
misc_feature
/note="wgs_contig"

ORIGIN
Query Match 5.8%; Score 90.2; DB 2; Length 217688;
Best Local Similarity 84.9%; Pred. No. 8.8e-08;
Matches 101; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 522 CAGAGAGATGAGACGAGTGTGCGGTGCGGAGCGCGGAGAGCTCATCTGCTGTGAC 581
DB 122922 CAGAGAGATGAGATGCGGTGCGGTGCGATGAGCGAGGCGAGCTCATCTGCTGTGAC 122863
QY 582 GGCTGCGCTCGGCGCTTCCACCGGCGCTGCTGCTCCCGCTCCGCGGAGATCCCCAG 640
DB 122862 GGCTGCGCTCGGCGCTTCCACCGGCGCTGCTGCTCCCGCTTGTGAGAGATCCCCAG 122804

RESULT 41
AC109383 283951 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-48B22, *** SEQUENCING IN PROGRESS
AC109383
AC109383.5 GI:25006726
HTG: HTGS_PHASE1, HTGS_DRAFT, HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 283951)
Muzny,D,Matie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alstbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, K., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazo, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,

```

```

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hune, J., Idelbit, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, S., Nguyen, N., Norris, S.,
Nwackeleme, O., Okunolu, G., Olarinmaga, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plommer, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
Rivers, C., Rodkey, T., Rojals, A., Rose, M., Rose, R., Ruiz, S. U.,
Sandera, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shwartsbeyn, A., Sisom, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soza, U.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tinney, L., Trejos, Z., Uman, K.,
Vais, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, F.,
Williams, G., Willson, R., Wleczek, R., Woodson, H., Wolley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Weidenhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 283951)
Worley, K. C.
Direct Submission
Submitted (04-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 283951)
Rat Genome Sequencing Consortium.
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23677016.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
genome sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPWN
Center clone name: CH230-48B22
----- Summary Statistics

```

Assembly program: Phrap; version 0.990329  
Consensus quality: 220070 bases at least Q40  
Consensus quality: 223733 bases at least Q30  
Consensus quality: 226333 bases at least Q20  
Estimated insert size: 228206; sum-of-coverage estimation  
Quality coverage: 6x in Q20 bases; sum-of-coverage estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.bgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 17097: contig of 17097 bp in length  
\* 17098 17197: gap of unknown length  
\* 17198 93015: contig of 75818 bp in length  
\* 93016 93115: gap of unknown length  
\* 93116 162410: contig of 69295 bp in length  
\* 162411 162510: gap of unknown length  
\* 162511 281691: contig of 119181 bp in length  
\* 281692 281791: gap of unknown length  
\* 281792 283951: contig of 2160 bp in length.  
FEATURES  
source 1.283951 Location/Qualifiers  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-48B22"  
803..35510  
/note="clone boundary  
clone\_end:Sp6  
site:  
end\_sequence:BH362533"  
33035..44376  
/note="clone boundary  
clone\_end:Sp6  
site:  
end\_sequence:BH362533"  
88940..90031  
/note="wgs end extension  
clone\_end:Sp6"  
91231..93015  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
ORIGIN  
Query Match 5.8%; Score 90.2; DB 2; Length 283951;  
Best Local Similarity 84.9%; Pred. No. 8.4e-08;  
Matches 101; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
QY 522 CAGAGAAATGAGAGCAGATGTGCTGTGTGCGGAGCGGGAGGCTCATCTGCTGTGAC 581  
|||||  
DB 22989 CAGAGAAACGAGAGATGATGCGCTGTGTGCATGACGAGGAGACTCATCTGCTGTGAC 23048  
|||||  
QY 582 GCGTGGCTTGGGCGCTTCCACCTGCGCTGTGCTCCCTCCGCTCCGGAGATCCCCAG 640  
|||||  
DB 23049 GGCTGTCCCGCGCTTCCACCTGCGCTGTGCTCCCGACCTCTGACGAGATCCCCAG 23107  
|||||  
RESULT 42  
LOCUS BV163805 558 bp DNA linear STS 15-MAY-2004  
DEFINITION RPAWMSB00036732 Roche Palo Alto Mus musculus STS genomic, sequence  
tagged site.  
ACCESSION BV163805  
VERSION BV163805.1 GI:47270013  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 558)  
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,  
McPherson,J.D., Foerzler,D. and Peltz,G.  
TITLE  
Mus musculus STS  
JOURNAL  
Unpublished (2003)  
COMMENT  
Contact: Jonathan Usuka  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: jonathan.usuka@roche.com  
Primer A: No primer submitted with this STS  
Primer B: No primer submitted with this STS.  
Location/Qualifiers  
1..558  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/map="10-7927-746-CBA01111727.1.1.11206"  
/clone\_11b="Roche Palo Alto"  
/note="SNPs developed from assay sequences derived from 15  
different strains of mice (as of October 1, 2003). Those  
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/OSHJ,  
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,  
MRL/MpJ, NZB/BinJ, NZW/LacJ, SPRET/Ei."  
<1..>558  
ORIGIN  
STS  
Query Match 5.3%; Score 82.6; DB 11; Length 558;  
Best Local Similarity 77.5%; Pred. No. 7e-06;  
Matches 100; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 918 CAGCAGAACTGCTCTGCTGCGGCTGCGGAGATGTAACGACGTGCTG 977  
|||||  
DB 392 CTGCAGGGGTCCAGCAGCAAGCGCGCATGCTGTGTGGAGATGACACGAGGTGTTG 451  
|||||  
QY 978 CCGTGTACTCATCTGCGCGCTGCTTCCATCTGACCTGCTCCAGCCGACCTTC 1037  
|||||  
DB 452 CGGTGTGACACTGTGCGCTGCTTCCATCTGACCTGCTCCAGCCGCGCGCC 511  
|||||  
QY 1038 CGGCCCGGG 1046  
|||||  
DB 512 CGGCCCGGG 520  
|||||  
RESULT 43  
LOCUS BV089350 574 bp DNA linear STS 15-OCT-2003  
DEFINITION RPAWMSB00001246 Roche Palo Alto Mus musculus STS genomic, sequence  
tagged site.  
ACCESSION BV089350  
VERSION BV089350.1 GI:37666829  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 574)  
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,  
McPherson,J.D., Foerzler,D. and Peltz,G.  
TITLE  
Mus musculus STS  
JOURNAL  
Unpublished (2003)  
COMMENT  
Contact: Jonathan Usuka  
Roche Palo Alto Genetics and Genomics Department  
Roche Palo Alto  
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA  
Tel: 6508555807  
Email: jonathan.usuka@roche.com  
Primer A: No primer submitted

## Primer B: No primer submitted.

FEATURES  
Location/Qualifiers

1..574

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/map="10-7927-7476-CANA0111727.1.1.11206"

/clone\_lib="Roche Palo Alto"

/note="SNPs developed from assay sequences derived from 15 different strains-of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D2-H2/osnJ, BALB/cBy, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J, MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei.."

<1..>574

STS

Query Match  
Best Local Similarity 77.5%; Score 82.6; DB 11, Length 574;  
Matches 100; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 918 CAGCAGACCTGCTCTGCTGCGGCTTGGCGGAGTGTGAGAGTGTGCTG 977

DB 408 CTGCAAGGTCACGACCAAGCGCGATGCTGTGTGCGATGCGACCGAGGTGTG 467

QY 978 CGGTGACTCACTGCGCGCTGCTTCCACTGCGCTGCTTCCAGCGGACCTTC 1037

DB 468 CGGTGACACACTGCTGCGCTTCCACTGCGCTGCTTCCAGCGGCGCGCG 527

QY 1038 CGGCCCGGG 1046

DB 528 CGGCCCGGG 536

RESULT 44

LOCUS BV088992 593 bp DNA linear STS 15-OCT-2003

DEFINITION RPAMMSRG0000888 Roche Palo Alto Mus musculus STS genomic, sequence

Accession BV088992 GI:37666471

VERSION BV088992.1

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE Uenaka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.

AUTORS Mus musculus SNPs

TITLE Unpublished (2003)

JOURNAL

COMMENT

Contact: Jonathan Uenaka

Roche Palo Alto Genetics and Genomics Department

3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA

Tel: 6508555807

Email: Jonathan.Uenaka@roche.com

Primer A: No primer submitted

Primer B: No primer submitted

Location/Qualifiers

1..593

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/map="10-7948-7452-CANA0111727.1.1.11206"

/clone\_lib="Roche Palo Alto"

/note="SNPs developed from assay sequences derived from 15 different strains-of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D2-H2/osnJ, BALB/cBy, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J, MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei.."

<1..>593

STS

ORIGIN

Query Match  
Best Local Similarity 77.5%; Score 82.6; DB 11, Length 593;  
Matches 100; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 918 CAGCAGACCTGCTCTGCTGCGGCTTGGCGGAGTGTGAGAGTGTGCTG 977

DB 403 CTGCAAGGTCACGACCAAGCGCGATGCTGTGTGCGATGCGACCGAGGTGTG 462

QY 978 CGGTGACTCACTGCGCGCTGCTTCCACTGCGCTGCTTCCAGCGGACCTTC 1037

DB 463 CGGTGACACACTGCTGCGCTTCCACTGCGCTGCTTCCAGCGGCGCGCG 522

QY 1038 CGGCCCGGG 1046

DB 523 CGGCCCGGG 531

RESULT 45

LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002

DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.

Accession AF429315

VERSION AF429315.1 GI:17646244

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 125020)

Ingber, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Huang, H.S.,

Potter, N.T., Ross, C.A. and Margolis, R.L.

A repeat expansion in the gene encoding junctophilin-3 is

associated with Huntington disease-like 2

Nat. Genet. 29 (4), 377-378 (2001)

PubMed 11694876

JOURNAL

MEDLINE

TITLE

AUTORS

REFERENCE

PUBMED

1 (bases 1 to 125020)

Holmes, S.E., Ingber, S.E., Rosenblatt, A., Callahan, C., Huang, H.S.,

Potter, N.T., Ross, C.A. and Margolis, R.L.

Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

Location/Qualifiers

1..125020

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S520 and WI-12410"

/note="Isolated from a patient with Huntington's

Disease-Like 2 (HDL2)"

/rpt\_type=tandem

/rpt\_unit="ctg"

/complement(36507..36887)

/gene="JPH3"

/note="synonym: JPH3"

/complement(36507..36887)

/product="JPH3"

/complement(36507..36887)

/note="component of the junctional complex between plasma

membrane and endoplasmic reticulum"

/codon\_start=1

/product="junctophilin 3"

/protein\_id="AA140941.1"

/db\_xref="GI:17646245"

/translation="MSSGGRFNPDDGSGYCGWEDGKAHGVCTGPKQGGYTGSGS

HFPELVGVYTWGNTYOGTMAQGGIGLSKQKMWYKGGWTHGFKR YGVRECAQ

NKAKYEGTWSNGLDQYGTETYSDC"

ORIGIN

Query Match	5.0%; Score 77.2; DB 9; Length 125020;
Best Local Similarity	12.2%; Pred. No. 3.3e-05;
Matches	116; Conservative 421; Mismatches 406; Indels 10; Gaps 4;
Qy	588 CCTCGGCGCTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATATCCCACTGAGGACC 647
Db	16887 MCVYSKSMMSMBMSVSVSMHMSHMSBSCMBKMTMSCCMSMYKSMMSMSMGWC 16946
Qy	648 TTAGAGTGTCTCCAGCTGCTGACGAGCAACATCCAGAGTGTGACGCCCGGAGAGAG 707
Db	16947 WGRRRKGMWYSRGMRSKSMRYTGSKKMRSNMCTSCVASMCMCWCSCCMRSCC 17006
Qy	708 CCCCCCGCAGAGGACCCCGGAGAGCCCGCTCCCGCGGCTTATGCTGAGCGGAG 767
Db	17007 CMCRSTCCCMRYCACCTWMSSTWYMSATMSRSTWYKSMKMSMCRSRSKSKRG 17066
Qy	768 GAGAGGTAAAGAGTCCACCTGGGGAACCCCTAGCCGAGATGACAGACTTGTCTAC 827
Db	17067 CGSMGKGKGSYGRKTRKSRGKMGAKMYWYRRSRKMYSSKGYCM--YCMC 17123
Qy	828 AAGCACTGCGCGCTCCGCTTGTGACGCCCGCTCCAGTGTGACCTCTCGCCCTG 887
Db	17124 WGRGCTCSCMTRSAMCSYCAKCKSMCYSCYGMKMGYYAASYRGSMSKXCMRG 17183
Qy	888 CACCCCTACTGTGTGGGTCTGAGGTCAGACAGACCTGCTCCTGTCGCGCTGCG 947
Db	17184 STYSTRCGCCCTTTTCCCCCNAATGGGAAAGTTTNCNKTYSTRKNGCAMKXNN 17243
Qy	948 GGGGTGTGGAGAGATGTAAGACGTGTGCGGTGTACTACGCGCCGCTGCTTCCAC 1007
Db	17244 YNNMSWRSCBAGMSCTK-----YKSMTMSMASYCMCMSTYCSMRSMASRMSYV 17298
Qy	1008 TGGCGCTGCACCTTCCACGCGGACCTCCCGGCGGAGCGGCTGCGCTGCAATCC 1067
Db	17299 KMKSSWRMSCYMKCCWCMKMCYCMRSMRSMSYMYASMSSSRGCTCTRCYCMS 17358
Qy	1068 TGCTCAGAGAGACGTGACCCAGCCCTGTGAGAGGGGTGAGCCCGCCAGCCCGC 1126
Db	17359 KSCYTSYVMMRSKMKMKMSKMSKMSMSASRSSCYKCYKMSKMSMSKRYRAGCM 17418
Qy	1127 CCTGCGCCCTGGCGCTGCCAAGATACACTGCACTCAGACCGCGCTTGACAGGA 1186
Db	17419 MKGCTMYRMYRCSMRBRWMSKAMYRMRMRMGKRGAMMMCMKCYRMSRCMMMSYCA 17478
Qy	1187 TGAACGTGAGTCCCTCTGAGGAGACACCTTCGANTG-GCATCCGTGACGTGGCCATCC 1245
Db	17479 SCRSKMSMSGTWKCASYCMSSYCTCWTCTSYSTRCTCWCCKMSSTYKSKSSSSMS 17538
Qy	1246 AAGAGATGAGCCCGTCCGCGGCGCCCTTCCCTCCAGACCCAGATGCGCGGACATGA 1305
Db	17539 SYKGRKSYSMCTRGAMSCMRRCYRGMASMSMAASRRARAGRSWGSKXMMWG 17598
Qy	1306 GCTGTATGAGAGGTGCTGAGAAGACACTCTTCTCTCACTGCTGGAAGCGCGCGG 1365
Db	17599 GMMRSKYYYCTGRMMMTYCCCMRRRSYMYRSMAMGMRMSGMRMMSASRRRC 17658
Qy	1366 TGGAGATCAAGAAGGAGCAGCGCCACTCTGTCACTGCTCGGCTGTAAACACTCTGTG 1425
Db	17659 KMSRSRSCSRKMRGSGSCMSKMWGSKRSASCKGSRMRRSKRSYRRRG 17718
Qy	1426 TTTCTGAGACACACCATCATGTGCTGAAATTAACCTCGCCCACTTCTTAATC 1485
Db	17719 KXRSMTKSKGSKKSKCWMRSGMTSSCYYSASCMMSMSSCMCCMMMRACACC 17778
Qy	1486 TGAAGTCCCCGGAGCCTCTCTTGCGCTGTGACTACTATAAATATATAAAA 1538
Db	17779 SSMSTSMYRCCMGKMSYSTCGTCKCTGYKKSRTWYMYMSWTSMYMR 17831

RESULT 46  
AF429315/c 125020 bp DNA linear PRI 18-JAN-2002  
LOCUS AF429315

DEFINITION	Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION	AF429315
VERSION	AF429315.1 GI:17646244
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS	1 (bases 1 to 125020) Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll,Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L. A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2
TITLE	Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL	21583737
MEDLINE	11694876
REFERENCE	2 (bases 1 to 125020) Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
AUTHORS	Direct Submission
TITLE	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
JOURNAL	Location/Qualifiers
FEATURES	1..125020
source	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="16" /map="16q24.3; between D16S520 and WI-12410" /note="Isolated from a patient with Huntington's Disease-like 2 (HDL2)" complement (35581..35746) /rpt_type=tandem /rpt_unit="ctg" complement (36507..36887) /gene="JPH3" /note="synonym: Jp3" complement (36507..36887) /gene="JPH3" complement (36507..36887) /product="Junctophilin 3" /note="component of the junctional complex between plasma membrane and endoplasmic reticulum" /codon_start=1 /product="Junctophilin 3" /protein_id="AA140941.1" /db_xref="GI:17646245" /translation="MSSGGRFNPDDGSGYCGWEDGKAHGVTGTPKQGGEYTGSMV HGFVLYGVYTPSGNTYQGTWAGKRGHIGLESKGMVYKGEWTHGFKGRYGVRECAAG NGAKYEGTSMNLDQGYGTETYSQG"
CDs	
ORIGIN	
Query Match	4.8%; Score 73.6; DB 9; Length 125020;
Best Local Similarity	11.3%; Pred. No. 0.00016;
Matches	113; Conservative 452; Mismatches 422; Indels 10; Gaps 5;
Qy	209 GGCCCGACCTGCGCTGAGAGAGGTTTCATGTGTTGCTACAGTCCGGGCGCCCTGG 268
Db	17859 KGCMTGTSKSKGKKTAYASRKRGKGTYYKRMASMSKMMRWASMRCAAGMGAAG 17800
Qy	269 AAGCAGACACCTGCAAGAACCAGGTTTCTTCCCAATAGGATGAGCCCGGGGGGTG 328
Db	17799 ASBSMKCMCGYRKASGSKSNGGTGYMKKGGGKSGMSKXMKSSSTRRSASAKSC 17740
Qy	329 CTGTTGAGACAG-ATGATGAGGCAACAGTGTGACGAGGCAAGATTGACGCCCTGGA 387
Db	17739 SYWGMSCMCMSSAKSYTMCYCYMYRMSYSWYKCYSGCMGSTSYSCCMKMS 17680
Qy	388 GCATGGAGACGAGACACTGGGAGTTACAGTTACCAAGATGCTCTGCGGAGGCT 447
Db	17679 WSGCYKCMKYSGMSSYSTSMGYSTSCCKTKCSWSMYKCTKSRYRKSSYYTWGGG 17620

```

Oy 448 GTTTGGGAAGAGGAGGCTCTGACAGAGGCTGTCACCCAGCCAGTCTGATGGAGC 507
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17619 KAKAKKYYCAGRRRRSSYMKCCAKMMMSYCMYSCTMYYSKSTYKSSTCYRGGYWG 17560
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 508 TCTCTGCGGTCGACGAAGATGAGAGAGGTCGCGGTCGAGGAGGAGGAGG 567
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17559 SKTCTYAGGKSNMYCMMRSSKSSSSWSMSMASSMCMGAGYRRSKSAGAGRRSS 17500
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 568 TCATCTGCTGTGACAGGCTGCGCTCGAGGCTTCCACCTGCGCTGCTCCCTCCGCTCC 627
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17499 KGRKSTG--KRAKSKTGSYGSYGRSMKKKGYSKYRSRMKMKKTKCYCMWYKRYKRT 17442
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 628 GGGAGATCCCACTGAGGAGCTGAGGCTGCTCCAGCTGCTCCAGCAACAGTCCAGAGG 687
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17441 SMCMMWYMKSWGYKRYKRCCKMKKGGCTGYMGSSKKSgySMRGMSSYSTCWMSSCWG 17382
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 688 TGCAGCCCGGAGAGAGAGCCCGGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17381 YSMKMKMTKYSYKRRSRMSMSKMGYAGRGCTSSMSMSTRKRSKCYKSYKKG 17322
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 748 CCGAGCTTAGTGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17321 RGRMKMGKMGKRSKYMSMKMKRSMSKYSTKYSGRKSKGMRGRTKAKSSMR 17262
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 808 -TGGAACAGACTTGTCTCAACAAGACCTGCGGCTCCGCTTCTGAGAGAGAGAGAGAG 866
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17261 MGSKCTTYSYNNRRNNRMKGTCTNNYRRSRMMNMAAAGCTTCCANTMGAG 17202
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 867 GGTCTGAGACTCTCGGCTGACACCCCTGACCTGCTGAGGCTGAGAGAGAGAGAGAG 926
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17201 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 17142
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 927 CTGAGCTCTGCTGCTGCGGCTGCGGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAG 986
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17141 GKTSYBAKSGRGCYCMWGRKGRKCMSSRYRRKMMYKSYRRKMMTCKMKCYSMYMWY 17082
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 987 CACTGCGGCTGCTTCCACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1046
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17081 CSBMCCKCKSCCGYSGMSMSYSYSGKXSMKMSYMSYSYKSS--TSKMRSSKRG 17024
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1047 ACAGGCTGCTGCTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17023 MGTGAYKKGGRYKGGGSGYKGGGSGWGGKSGKSTGSSAGKSS---SYKMSCCARYK 16968
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1107 CTGGCCCCAGCCCCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1166
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16967 SMSYKCYRRMMWCMYYCWMGKCMSSCWMSSMRKKSCKGSAKGMVMDKGSVTDKSD 16908
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1167 GAGCCGCTCTGACAGGATGACCTGAGTCCCTTC 1203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16907 KMSRSBSKVKSKMSMRGKYYCTDSMKMYTGC 16871
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 47
BX649390 184594 bp DNA linear VRT 11-FEB-2004
LOCUS Zebrafish DNA sequence from clone DKEY-261L2 in linkage group 2,
DEFINITION complete sequence.
ACCESSION BX649390
VERSION BX649390
KEYWORDS HTG:42538836
SOURCE Danto rerio (zebrafish)
ORGANISM Danto rerio

```

```

REFERENCE
AUTHORS Johnson, C.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
COMMENT On Feb 11, 2004 this sequence version replaced gi:4251008.

```

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmark.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml)

VECTOR: pIndigoBAC-5.

```

FEATURES
    source
        1..184594
            /organism="Danto rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-261L2"
            /clone_1ib="Dantokey"

```

#### ORIGIN

```

Query Match 4.5%; Score 70; DB 5; Length 184594;
Best Local Similarity 74.6%; Pred. No. 0.00077;
Matches 88; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

```

Oy 523 AGAAGATGAGAGAGAGTGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78081 AGCACAATATGATGATGTGACAGTGTGAAGACGCGCGAGCATCTGTGTGATG 78022
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 583 GCTGCCCTGGGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 640
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78021 GGTGTCTGTGTCTTCACTCTGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 77964
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 48
AK127046 4966 bp mRNA linear PRI 19-FEB-2004
LOCUS Homo sapiens cDNA FL45103 f18, clone BRMH3032571, moderately
DEFINITION similar to Chromodomain helicase-DNA-binding protein 4.
ACCESSION AK127046
VERSION AK127046.1 GI:34533779
KEYWORDS oligo capding; f18 (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,

```



Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuna, M., Murakawa, K., Kanenori, K., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T., NEDO human cDNA sequencing project Unpublished 2 (bases 1 to 4966) Isogai, T. and Yamamoto, J. Direct Submission Submitted (15-JUL-2003) Takao Isogai, Flu Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) ; cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.

COMMENT

JOURNAL  
TITLE  
AUTHORS  
JOURNAL

FEATURES  
source

CDS

Location/Qualifiers  
1. .4966  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="BRAMH3032571"  
/issue\_type="brain"  
/clone\_lib="BRAMH3"  
/note="cloning vector: pME18SFL3"  
1. .3678  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAC68602.1"  
/db\_xref="GI:34533780"  
/translation="MRGPVTEEBELPRLFAEMENEDENSEEEDGGLAEPDFPVEP  
VLPKKKKRKKLKENKCGKRRKKSNDLSENEDELREKSESGSDVSPMKKKK  
LKDKEKAKRRKKDEDDNDGCLKEPKSGGLMAEWGLDQVLYPSEBYHTLTN  
YAFSGQPLRLAKKPKTPMSKMTVLAKEKREBSANNPFKSSAAAAAVALAVE  
TVTISPLAVSPQVPVIRAKTKESGPGVRAKIKSGDGGKGGKGTAGAKF  
RFGISNNRKKSSSEDEBESDPSAISHASVRSBSAALGKSKRRKKRID  
GGVETIDHDQYCEVCOQGEIILCDTCEPAVHLVCLDPELEKAPBGKSCPHCEKGI  
QMEPKDDDEBEGCEEBEDDMEFCRCVCKDGEILCCDAPSSYHLCNLPLEI  
PGEWMLCPRCCTPPLKGVORILHMRTEPPAPFMVGLPGPVESELPPKPLEGIP  
REPRVYKALSYWHCSWYKELQELVHTMYRYQRKNDMPPEPPDYGSGEDGSE  
KRNKQDPLAKMEERFYRGIRPEMMHRIILNHSFDKGDVHTLTKMDLPDQCTW  
ELDIDIPYVDNLQAAYWGHRELMLGEDTGLPRLKLGKRLDKQEKRPDTPVDP  
TVKFDKQPVYIDSTGTLHPYQLEGLNMLRFSWAQGTDTLADEMGLGTVQTVLY  
SLYKESGKPYLVASPLSTIINMREPEMMAADPFVVTYTGDKSRSVIRNEFSFE  
DNAIRSGKVPFMKKEVOIKFVHLTSYELITIDQALIGSIEMACLVDVEARHLKNO  
SKPRVLNSYKIDYKLLTGTPLONNBELFHLNPLFTRPRNNLEGPLEPRADISKE  
DOIKUHLDLGPHMLRLKADVFKNMPAKTELIVRVLSQMKYTKFILTFRFBLN  
SKGGNQVSLNIMDLKCCNHPYLPVAALVAVLPNGSYDSSLVSSGKMLIQ  
KMLKRLRDEGHRVLPFQMTKMLDLLEDFLEBYGYKERIDG1TGLAQEADIRFNA  
PQAQOFCPLSTRAGGLINTLATDVTLIYDSMPNPHNDIOAFSAHRIIGOTKWKESW  
ATMWRKTGPHRSQSGSPMKSRBPQPMWIMRWVITSTRAQPARREBSGQRRRRRPP  
RSSCRERRCFLLRRRSWTSMWS"

ORIGIN

Query Match 4.2%; Score 65; DB 9; Length 4966;  
Best Local Similarity 62.7%; Pred. No. 0.013;  
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 532 AGGACAGTGTGCGCGTGTGGGACGGCGGGAGCTCATCTGCTGACGGCTGCCCTC 591  
DB 1028 AGGATTACTGTGAGTGTGCGACAGGGTGGGAGATCATCTGCTGACACCTGCCCGCA 1087  
QY 592 GGGCCTTCCACCGGGCTGCGTCCCGCTCCGGGAGATCCCGACAGTGGGAGACCTGGA 651  
DB 1088 GGGCCTTCCACCTGCGTATGCTTGAGACCCAGAGCTGGAAGAGGCTCCCGAGGGCAAGTGA 1147  
QY 652 GGTGCTCAGCTGCTGTCAGGCAACAGTCCAGAGAGGTGACG 922

Db 1148 GCTGCCCCCATCTGTGAGAAAGAGGGGATCCATCGGAGCGC 1188  
RESULT 49  
AF425231  
LOCUS  
DEFINITION  
Homo sapiens chromodomain helicase DNA binding protein 5 (CHD5)  
mRNA, complete cde.  
AF425231.1 GI:19773959  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL  
Submitted (27-SEP-2001) Oncology, Children's Hospital of Philadelphia, 3615 Civic Center Blvd., Philadelphia, PA 19104-4318, USA

FEATURES  
source

gene  
CDS

Location/Qualifiers  
1. .9646  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="1p36.3"  
1. .9646  
/gene="CHD5"  
101. .5965  
/gene="CHD5"  
/codon\_start=1  
/product="chromodomain helicase DNA binding protein 5"  
/protein\_id="AAJ98962.1"  
/db\_xref="GI:19773960"  
/translation="MRGPVTEEBELPRLFAEMENEDENSEEEDGGLAEPDFPVEP  
VLPKKKKRKKLKENKCGKRRKKSNDLSENEDELREKSESGSDVSPMKKKK  
LKDKEKAKRRKKDEDDNDGCLKEPKSGGLMAEWGLDQVLYPSEBYHTLTN  
YAFSGQPLRLAKKPKTPMSKMTVLAKEKREBSANNPFKSSAAAAAVALAVE  
TVTISPLAVSPQVPVIRAKTKESGPGVRAKIKSGDGGKGGKGTAGAKF  
RFGISNNRKKSSSEDEBESDPSAISHASVRSBSAALGKSKRRKKRID  
GGVETIDHDQYCEVCOQGEIILCDTCEPAVHLVCLDPELEKAPBGKSCPHCEKGI  
QMEPKDDDEBEGCEEBEDDMEFCRCVCKDGEILCCDAPSSYHLCNLPLEI  
PGEWMLCPRCCTPPLKGVORILHMRTEPPAPFMVGLPGPVESELPPKPLEGIP  
REPRVYKALSYWHCSWYKELQELVHTMYRYQRKNDMPPEPPDYGSGEDGSE  
KRNKQDPLAKMEERFYRGIRPEMMHRIILNHSFDKGDVHTLTKMDLPDQCTW  
ELDIDIPYVDNLQAAYWGHRELMLGEDTGLPRLKLGKRLDKQEKRPDTPVDP  
TVKFDKQPVYIDSTGTLHPYQLEGLNMLRFSWAQGTDTLADEMGLGTVQTVLY  
SLYKESGKPYLVASPLSTIINMREPEMMAADPFVVTYTGDKSRSVIRNEFSFE  
DNAIRSGKVPFMKKEVOIKFVHLTSYELITIDQALIGSIEMACLVDVEARHLKNO  
SKPRVLNSYKIDYKLLTGTPLONNBELFHLNPLFTRPRNNLEGPLEPRADISKE  
DOIKUHLDLGPHMLRLKADVFKNMPAKTELIVRVLSQMKYTKFILTFRFBLN  
SKGGNQVSLNIMDLKCCNHPYLPVAALVAVLPNGSYDSSLVSSGKMLIQ  
KMLKRLRDEGHRVLPFQMTKMLDLLEDFLEBYGYKERIDG1TGLAQEADIRFNA  
PQAQOFCPLSTRAGGLINTLATDVTLIYDSMPNPHNDIOAFSAHRIIGOTKWKESW  
ATMWRKTGPHRSQSGSPMKSRBPQPMWIMRWVITSTRAQPARREBSGQRRRRRPP  
RSSCRERRCFLLRRRSWTSMWS"



## ORIGIN

MAGNKPANALVHLKQLEBLLSDMKADVTBRLPATIARIPIVAVRLOMSEBRLISRLA  
NEAPEPPEQOVAQOQ"

Query Match 4.1%; Score 63.2; DB 10; Length 6438;  
Best Local Similarity 67.4%; Pred. No. 0.028;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGCAGTGTCCGCTGTGTGCGGACGGCGGAGCTCATCTGTGTGACGGCTGCCCTCG 592  
DB 1503 GGAATTTCTGTGCGGTGTGACGAGACGGCGGAGCTCTGTGTGTGACACATGCCCTTC 1562  
QY 593 GGCCTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652  
DB 1563 TTCTTACCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1622  
QY 653 GTGCTCCAGCTG 664  
DB 1623 CTGTCCCGCTG 1634

RESULT 51  
AC122760 67573 bp DNA linear HTG 15-JUN-2002  
LOCUS Mus musculus clone RP24-489C24, LOW-PASS SEQUENCE SAMPLING.  
AC122760  
AC122760.2 GI:21427758  
VERSION HTG, HTGS\_PHASE0.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Birren, B., Linton, L., Nussbaum, C. and Lander, R.  
TITLE Mus musculus, clone RP24-489C24  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 67573)  
AUTHORS

REFERENCE  
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Keils, C., Labocque, K.,  
Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,  
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicot, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (25-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 67573)  
REFERENCE  
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Keils, C., Labocque, K.,  
Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K.,

TITLE  
JOURNAL

## COMMENT

Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,  
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicot, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 15, 2002 this sequence version replaced gi:21206372.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L26032  
Center clone name: 489\_C\_24  
-----

\* NOTE: This record contains 84 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 711: contig of 711 bp in length  
\* 712 811: gap of 100 bp  
\* 812 1523: contig of 712 bp in length  
\* 1524 1623: gap of 100 bp  
\* 1624 2341: contig of 718 bp in length  
\* 2342 2441: gap of 100 bp  
\* 2442 3148: contig of 707 bp in length  
\* 3149 3248: gap of 100 bp  
\* 3249 3962: contig of 714 bp in length  
\* 3963 4062: gap of 100 bp  
\* 4063 4778: contig of 716 bp in length  
\* 4779 4878: gap of 100 bp  
\* 4879 5593: contig of 715 bp in length  
\* 5594 5693: gap of 100 bp  
\* 5694 6406: contig of 713 bp in length  
\* 6407 6506: gap of 100 bp  
\* 6507 7225: contig of 719 bp in length  
\* 7226 7325: gap of 100 bp  
\* 7326 8023: contig of 698 bp in length  
\* 8024 8123: gap of 100 bp  
\* 8124 8820: contig of 697 bp in length  
\* 8821 8920: gap of 100 bp  
\* 8921 9619: contig of 699 bp in length  
\* 9620 9719: gap of 100 bp  
\* 9720 10417: contig of 698 bp in length  
\* 10418 10517: gap of 100 bp  
\* 10518 11224: contig of 707 bp in length  
\* 11225 11324: gap of 100 bp  
\* 11325 12037: contig of 713 bp in length  
\* 12038 12137: gap of 100 bp  
\* 12138 12837: contig of 700 bp in length  
\* 12838 12937: gap of 100 bp  
\* 12938 13661: contig of 724 bp in length  
\* 13662 13761: gap of 100 bp

```

13762 14487: contig of 726 bp in length
14488 14587: gap of 100 bp in length
14588 15291: contig of 704 bp in length
15292 15391: gap of 100 bp in length
15392 16079: contig of 688 bp in length
16080 16179: gap of 100 bp in length
16180 16877: contig of 698 bp in length
16878 16977: gap of 100 bp in length
16978 17655: contig of 678 bp in length
17656 17755: gap of 100 bp in length
17756 18470: contig of 715 bp in length
18471 18570: gap of 100 bp in length
18571 19236: contig of 666 bp in length
19237 19336: gap of 100 bp in length
19337 20042: contig of 706 bp in length
20043 20142: gap of 100 bp in length
20143 20842: contig of 700 bp in length
20843 20942: gap of 100 bp in length
20943 21660: contig of 718 bp in length
21661 21760: gap of 100 bp in length
21761 22475: contig of 715 bp in length
22476 22575: gap of 100 bp in length
22576 23281: contig of 706 bp in length
23282 23381: gap of 100 bp in length
23382 24097: contig of 716 bp in length
24098 24197: gap of 100 bp in length
24198 24898: contig of 701 bp in length
24899 24998: gap of 100 bp in length
24999 25689: contig of 691 bp in length
25690 25789: gap of 100 bp in length
25790 26489: contig of 700 bp in length
26490 26590: gap of 100 bp in length
26590 27288: contig of 699 bp in length
27289 27388: gap of 100 bp in length
27389 28077: contig of 689 bp in length
28078 28177: gap of 100 bp in length
28178 28881: contig of 704 bp in length
28882 28981: gap of 100 bp in length
28982 29692: contig of 711 bp in length
29693 29792: gap of 100 bp in length
29793 30513: contig of 721 bp in length
30514 30613: gap of 100 bp in length
30614 31334: contig of 721 bp in length
31335 31434: gap of 100 bp in length
31435 32135: contig of 701 bp in length
32136 32235: gap of 100 bp in length
32236 32931: contig of 696 bp in length
32932 33031: gap of 100 bp in length
33032 33742: contig of 711 bp in length
33743 33842: gap of 100 bp in length
33843 34542: contig of 700 bp in length
34543 34642: gap of 100 bp in length
34643 35354: contig of 712 bp in length
35355 35454: gap of 100 bp in length
35455 36140: contig of 686 bp in length
36141 36240: gap of 100 bp in length
36241 36953: contig of 713 bp in length
36954 37053: gap of 100 bp in length
37054 37747: contig of 694 bp in length
37748 37847: gap of 100 bp in length
37849 38564: contig of 717 bp in length
38565 38664: gap of 100 bp in length
38665 39372: contig of 708 bp in length
39373 39472: gap of 100 bp in length
39473 40178: contig of 706 bp in length
40179 40278: gap of 100 bp in length
40279 40987: contig of 709 bp in length
40988 41087: gap of 100 bp in length
41088 41809: contig of 722 bp in length
41810 41909: gap of 100 bp in length
41910 42601: contig of 692 bp in length
42602 42701: gap of 100 bp in length
42702 43395: contig of 694 bp in length

```

```

Query Match
Best Local Similarity 67.4%; Pred. No. 0.019; Length 67573;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGCCTGCTGTGCGGACGCGCGGAGCTCATCTGCTGACGCTCCCTCG 592
Db 12393 GGAATCTCTTCGGCTGTGCAAGGAGCGGGAGCTCCTGCTGTGACATGCTTC 12452
QY 593 GGCCTTCACCTGAGCTGCTTCCCTCCCTCGCGGAGATCCCACTGGAGCTGAG 652
Db 12453 TTCCTACCAATCCATGCTGCTTAACCCCGCTGCGAGATCCCAAGGGAATGCT 12512
QY 653 GTCCTCCAGCTG 664
Db 12513 CTGTCCCGCTG 12524

RESULT 52
AC134529
LOCUS
DEFINITION
AC134529 173909 bp DNA linear HTGS 05-MAR-2003
MUS musculus chromosome UNK clone RP24-503A2, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC134529
AC134529.3 GI:28850175
HTGS: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
MUS musculus (house mouse)
ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerogamathi; Muridae; Murinae; Mus.
1 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
Direct Submision
Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
Direct Submision
Submitted (05-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Mar 5, 2003 this sequence version replaced gi:28626883.

REFERENCE
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Unpublished
JOURNAL
TITILE
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submision
JOURNAL
Submitted (05-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project Information
Center project name: M_BB0503A02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: pUC19; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175465 bases at least Q40
Consensus quality: 176221 bases at least Q30
Consensus quality: 176527 bases at least Q20
NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1200: contig of 1200 bp in length
1201
1300: gap of unknown length
1301
6880: contig of 5580 bp in length
6881
6980: gap of unknown length
6981
38527: contig of 31547 bp in length
38528
38628: gap of unknown length
38628
93610: contig of 54983 bp in length
93611
93711: gap of unknown length
93711
173909: contig of 80199 bp in length.
Location/Qualifiers
1. 173909
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-503A2"
misc_feature
1. 1200
/note="assembly_name:Contig22"
misc_feature
1301. 6880
/note="assembly_name:Contig27"
misc_feature
6981. 38527
/note="assembly_name:Contig28"
misc_feature
38628. 93610
/note="assembly_name:Contig29"
misc_feature
93711. 173909
/note="assembly_name:Contig30"
ORIGIN
Query Match
Best Local Similarity 4.1%; Score 63.2; DB 2; Length 173909;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGAAGAGTGTGCGTGTGTGCGGAGCGGAGGAGCTATCTGTGTGACGGCTGCCCTCG 592
12809 GGAATCTGTGCGTGTGTGAGAGCGCGGAGGAGCTCTGTGTGTGACATGACCTTC 12868

QY 593 GGCCTTCACCTGAGCTGCTGTGCTCCCTCCGCTCCGCGAGATGCCAGATGAGACCTGAG 652
12869 TTCTTACACATCTCCACTGCTGAAACCCCGCTGCCAGAGATCCCAAACGCGCAATGCGCT 12928

QY 653 GTGCTCCAGCTG 664
12929 CTGTCCCGCTG 12940

Db
12929 CTGTCCCGCTG 12940

RESULT 53
BV089378/c 572 bp DNA linear STS 15-OCT-2003
DEFINITION RPAHMSB00001274 Roche Palo Alto Mus musculus STS genomic, sequence
tagged site.
ACCESSION BV089378
VERSION BV089378.1 GI:37666857
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 572)
Uuska,T., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Poernitzler,D. and Peltz,G.
MUS MUSCULUS SNPS
Unpublished (2003)
TITLE
JOURNAL
COMMENT
Contact: Jonathan Uuska
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Uuska@roche.com
Primer A: No primer submitted

```

```

Primer B: No primer submitted.
Location/Qualifiers
1. 572
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="10-9361-9932-CABA0111727.1.1.11206"
/clone.lib="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains-of mice (as of October 1, 2003). Those
strains include A/J, A/Heu,-129/Sv, AKR/J, B10.D2-H2/oshu,
BALB/cByJ, BALB/cJ, C57BL/6J, -CAST/Ei, DBA/2J,
MRL/MPJ, NZB/BinJ, NZW/LacJ, SPRET/Ei.-"
<1. .>572
ORIGIN
STS
Query Match
Best Local Similarity 4.1%; Score 62.8; DB 11; Length 572;
Matches 82; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 634 TCCCAAGTGGGACCTGAGAGTCTCAGCTGCTCAGACGACACAGTCCAGAGGTGACAC 693
359 TACACAGTGGCTCTGAGAGATCTCTGTGCTCCAGGCGAGAGTCCAAACGAACCTT 300

QY 694 CCCGGGACAGAGAGCCCGCGCCCGACAGAGCCACCCGTGAGAGACCCCGCTCCCCC 748
299 CCCAGCTGAGGTGTCCAGAGCCCGCGGAGCAGACCTGACAGAGACCCCGTATGCC 245

Db
299 CCCAGCTGAGGTGTCCAGAGCCCGCGGAGCAGACCTGACAGAGACCCCGTATGCC 245

RESULT 54
AC097256/c 248835 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-172B6, *** SEQUENCING IN PROGRESS
DEFINITION *** 5 unordered pieces.
ACCESSION AC097256
VERSION AC097256.6 GI:30521113
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 248835)
REFERENCE
1 Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Angeliebachi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benhammed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cre,A., D'Souza,L.,
Davilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huijck,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewala,L., Loulseged,H., Lozada,R.J., Lu,X., Ma,D.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawmney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,

```

Markaverich, C., Neale, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwankwem, C., Okwuonu, G., Olanpinla, A., Pal, S., Parks, K.,  
 Paternak, S., Paul, H., Perez, A., Perez, D., Pfannkuch, C.,  
 Plopper, F., Poldinger, A., Popovic, D., Primus, E., Pu, L., L.,  
 Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M., Reigh, R.,  
 Rella, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., J.,  
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbain, A., Sisson, I., Sitter, C.D., Smales, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steinle, M., Strong, R., Sutton, A., Swale, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trege, Z., Usmani, K.,  
 Valdes, R., Vase, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G., and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 248835)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-OCT-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 248835)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced g1:23100986.  
 The sequence in this assembly is a combination of BAC tagged reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GIRK  
 Center clone name: CH230-17256  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 212993 bases at least Q40  
 Consensus quality: 215360 bases at least Q30  
 Consensus quality: 217739 bases at least Q20  
 Estimated insert size: 225359; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a working draft sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 23011: contig of 23011 bp in length









KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kites, such as nucleic acid arrays, comprising a majority of human exons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 8343 06-SEP-2002;  
FEATURES  
source  
1. 6412  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 3.9%; Score 60; DB 6; Length 6412;  
Best Local Similarity 61.5%; Pred. No. 0.12;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACAGTGTGCGGTGTCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTC 591  
DB 1193 AGGACTATTGCGAGGTCGACGAGCGCGGTGATCATCTGTGATACCTGTCC 1252  
QY 592 GGGCCCTTCCACCTGCGCTGCTGCCCTCCCTCCGAGAGATCCCAAGTGGAGCTGGA 651  
DB 1253 GTGCTTACCAACAGTGTGCTGCTGATCCGACATGAGAGAGCTCCGAGGCGAAGTGA 1312  
QY 652 GGTGCTCCAGCTGCTGCGAGGCAACAGTCCAGAGG 687  
DB 1313 GGTGCCACACCTGCGAGAGGAGGATCCAGTGGG 1348

RESULT 63  
AX332238  
LOCUS AX332238 6417 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 2747 from Patent WO0194629.  
ACCESSION AX332238  
VERSION AX332238.1 GI:18122872  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
Hortigan, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL Patent: WO 0194629-A 2747 13-DEC-2001;  
FEATURES  
source  
1. 6417  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 3.9%; Score 60; DB 6; Length 6417;  
Best Local Similarity 61.5%; Pred. No. 0.12;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 652 GGTGCTCCAGCTGCTGCGAGGCAACAGTCCAGAGG 687  
DB 1318 GGTGCCACACCTGCGAGAGGAGGATCCAGTGGG 1353

RESULT 64  
HSM12218  
LOCUS HSM12218 6417 bp mRNA linear PRI 13-MAR-1996  
DEFINITION H. sapiens mRNA for 218kD M1-2 protein.  
ACCESSION X86691  
VERSION X86691.1 GI:1107695  
KEYWORDS helixase; M1-2 gene; M1-2 protein.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Gentile, E., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M. and  
Seelig, H.P.  
TITLE The major dermatomyositis-specific M1-2 autoantigen is a presumed  
JOURNAL helixase involved in transcriptional activation  
MEDLINE Arthritis Rheum. 38 (10), 1389-1399 (1995)  
PUBMED 7575689  
REFERENCE 2 (bases 1 to 6417)  
AUTHORS Seelig, H.P.  
TITLE Direct Submision  
JOURNAL Submitted (26-APR-1995) H.P. Seelig, Institute of Immunology &  
Molecular Genetics, Kriegstrasse 99, D-76133 Karlsruhe, FRG

FEATURES  
source  
1. 6417  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/map="p13"  
/cell\_line="Hela"  
/clone\_lib="lambda"  
1. 6417  
/gene="M1-2"  
90. 5828  
/gene="M1-2"  
/codon\_start=1  
/product="M1-2 protein"  
/protein\_id="CAA60384.1"  
/db\_xref="GI:1107696"  
/db\_xref="GOA:Q14839"  
/translation="MASCIGSPSPGASGSEEDMALNLSPPHPENEDPDL  
ETETKLLKKKKPKPRPKPKPKRKKERMLLRQDSDSGSPPEVEEVEALR  
SDSBSDYTPGKKKKKKLGPKKKKSKRKEEEDDDDDSKPKSAQLLEWGM  
BDIDVFSEEDYRLTNYKAFQFVRPLIAKNPIAVSKMMVVGAKNPFSTNNP  
KSSGASVAAAAAIVAVESWTVAPPPPVAVPIRKAKTEGKPNRRPKG  
SPRVDAKKPKPKVAPLKKLGGFGSKSSSSDDDDLVSDDDGINSVSDG  
STSRSSRRKLLRTTKKKKKGSEETAYNDGYTDYDXYCENQCGEIIIDTCRAY  
HNVCLDPMEKAPBGKMSCPHCEKGIWEAKEDSBESBIEIEGDDDEEDHME  
FCVCKDGEELCCDTPSSVHICLNLPPEIPNGEWLCPKCTCPALKGKQKLLI  
KMGQSPSPTPVPRPDADPNTPSPGLGRPERQPFVVKMGSYNHCWVSLQLEH  
COMFRNRYORKNDMPSPGDEGDESKRKNDPKFAEKEEFYRGIKPEMMI  
HRIINHSVDKKGHYVLIKMRDLPYDOASWSESDVEIDYDYLFGSYNNHRLMGE  
GRGKKLKKVLRKLERPEPTPNTPYKSERQPRVLDATGTTLHPYMEGINMLRFS  
MAQGTDTLADBMGLKTVQTVAPLYSLYKESHSKGFPLVAPLSTIINWEFEKMA  
IDMALVTVGDDSDRAIIRNEPSFEDNIRGKASPMKKEASVRFVLLTSYELIT  
IDMALVGSIDVACLIIVDAHRKNNQSKFRFLVNGYSQHKLLLTGTPLONNLFEH  
ILNFLTPERFHLIEGFLSEFADIAKEDIKKLHMDLGHMLRLKADVFKNPSTEL  
IYRVLSFMOKKYVYILTRNPEALNARGGQVGLNVNMDLKKSCNHPYLPVAM  
EAPKMDNGVGSALIRASGKLLLOKLLKXEGSHVILFSQWTKMLDLEDELEH  
EGYKTERIDGGITGMROBALDRFAPBAQPCPLSTRBAGLGINTLATVTVIYDS  
DNPENNDIQAPSRARHIGQNKVMYIRFVTRASVERITTVAKKVMILHLVRLVGLG  
SKTSGMSKQEDLDILKFTGEBLFDQEARDDGDNKDEGSDSVTHYDDAIEKRLDRN  
DETDETLQGNMEYLSFKAQYVVRREMGSEEEVEREIIKQESVDPDVEKLLRH  
HYEQOEDLARNLIGKRLIKOVNNDGSDQEDRDQDDSDQSDSYVASSEGDGDFD  
RSEAPRRPRRGLANDKDKPLPLPLAVGNGIIVLGNRGRKAFNLAINMYGNPQ

DAFTOMLVIRDRGSEKFKAYVSLFEMHLCBPADGAEFADGPREGSLRCHVLT  
RIGWMSLIRKQYOBFRHVNRGMSMPELAEVENEKMSOGSGSPKPTTSTGTCDTDPN  
TPAPVPADGDKIEENSJKEESIEGEKVEVSTLAPETIIECTQAPAPSEKXVVE  
PEGEKVEKAVKERTERPEBTEPGADAVKVEKSAIDLTPIVDEKKEEB  
KEWMLONGETPNDKOKKNIKORFENFADGTEFLHLMQNEERAAVTKTY  
EIMHRRWYMLAGIINNGYAMODIONPRVALINEPKGMNNGNFLEIKNFKLAR  
RKLLLEALVIEBOLRRAYLMSDEPSPMSALNTRPAVEVCGLSEHSHLSKEMAG  
MKPANAVALHKLKQLEELISDMKADVTPLPATIATIPVAVRLQNSERNILSRLANRA  
PEPTPOVAQOQ

## ORIGIN

Query Match 3.9%; Score 60; DB 9; Length 6417;  
Best Local Similarity 61.5%; Pred. No. 0.12;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGAGCTCATCTGTGTGAAGGCTGCTTC 591  
DB 1198 AGGACTATTGCGAGGTGTGCGGACGCGGGAGAGCTCATCTGTGTGTATACCTGTCCCC 1257  
QY 592 GGGCCCTTCACCTGGCTGCTGTCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCCTGGA 651  
DB 1258 GTGCTTACCAATGCTGCTGCTGATCCGACATGAGAGGCTCCCGAGGCGCAAGTGA 1317  
QY 652 GGTGCTCCAGCTGCTGCTGAGGCAACAGTCCAGGAGG 687  
DB 1318 GCTGCCACACTGCGAGAGGAAGGATCCAGTGGG 1353

RESULT 65  
AR338834 6475 bp DNA linear PAT 17-AUG-2003  
LOCUS AR338834  
DEFINITION Sequence 325 from patent US 6569662.  
ACCESSION AR338834  
VERSION AR338834.1 GI:33725691  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 6475)  
AUTHORS Tang,Y.T., Zhou,P. and Drmanac,R.T.  
TITLE Nucleic acids and polypeptides  
JOURNAL Patent: US 6569662-A 325 27-MAY-2003;  
FEATURES Location/Qualifiers  
1..6475  
source /organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

Query Match 3.9%; Score 60; DB 6; Length 6475;  
Best Local Similarity 61.5%; Pred. No. 0.12;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGAGCTCATCTGTGTGAAGGCTGCTTC 591  
DB 1327 AGGACTATTGCGAGGTGTGCGGACGCGGGAGAGCTCATCTGTGTGTATACCTGTCCCC 1386  
QY 592 GGGCCCTTCACCTGGCTGCTGTCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCCTGGA 651  
DB 1387 GTGCTTACCAATGCTGCTGCTGATCCGACATGAGAGGCTCCCGAGGCGCAAGTGA 1446  
QY 652 GGTGCTCCAGCTGCTGCTGAGGCAACAGTCCAGGAGG 687  
DB 1447 GCTGCCACACTGCGAGAGGAAGGATCCAGTGGG 1482

RESULT 66  
BC038596 6584 bp mRNA linear PRI 07-OCT-2003  
LOCUS BC038596  
DEFINITION Homo sapiens chromodomain helicase DNA binding protein 4, mRNA  
ACCESSION BC038596  
VERSION BC038596.1 GI:24047225  
KEYWORDS MGC.

SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Homo sapiens  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Euteheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 6584)

REFERENCE  
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Schert,T.E., Brownstein,M.J., Uedlin,T.B., Toshiyuki,S.,  
Carrinci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kerteman,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Boutard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywicki,M.I., Skalek,U., Smalins,D.E.,  
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22386257  
JOURNAL MEDLINE  
PUBMED 12477932  
TITLE 2 (bases 1 to 6584)  
AUTHORS Strausberg,R.  
JOURNAL Direct Submission  
Submitted (15-OCT-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@ncl.nih.gov](mailto:nisc_mgc@ncl.nih.gov)  
Ahter,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
Dieterich,N.L., Granite,S., Guan,X., Gupta,U., Haghighi,P.,  
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Larc,P., Legaspi,R.,  
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,  
McDowell,U., Pearson,R., Stantripp,S., Thomas,P.J., Touchman,J.W.,  
Tsungeon,C., Vogt,J.L., Walker,M.A., Wechetby,K.D., Wiggins,L.,  
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>  
Series: IRAK Plate: 78 Row: 0 Column: 18  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4557452.

## FEATURES

source

gene

1..6584  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:46187 IMAGE:5528023"  
/cissue\_type="Skin, melanotic melanoma."  
/clone\_id="NIH\_MGC\_72"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..6584  
/gene="CHD4"  
/note="synonym: M12-BETA"

[illegible]

Query Match	3.9%	Score 60	DB 9	Length 6594
Best Local Similarity	61.5%	Pred. No. 0.12		
Matches 96	Conservative 0	Mismatches 60	Indels 0	Gaps 0

Qy	532	AGACGAGTGTGCGCTGTGTCTGGACGCGCGGAGCTCATCTGTGTGACGCGTGCCTC	591
Db	1262	AGGACTATTGCGAGAGTGTGCGACGACGAGCGGTGATCATCTGTGTGATCATCTGTCCCC	1321
Qy	592	GGGCCTTTCACCTGCGCTGCTGTCCCTCGCTCGCGAGATCCCACTGAGGACCTCGA	651
Db	1322	GTGCTTACCAATGTGTGTGCTGTGATCCCGACATGAGAGAGGCTCCGAGGCGCAATGGA	1381
Qy	652	GGTGCTCAGCTGCTGCGAGGCAACAGTCCGAGAGG	687
Db	1382	GCTGCCACACTGCGAGAGAGGACATCCAGTGGG	1417

RESULT 67	
AL589792	
LOCUS	132592 bp DNA linear HTG 10-JUN-2001
DEFINITION	Homo sapiens chromosome 1 clone RP5-889P23 map p36.21-36.33, 4
ACCESSION	AL589792
VERSION	AL589792.1 GI:13277194
KEYWORDS	HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE	1
AUTHORS	Lloyd, C.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
	requests: clonerequests@sanger.ac.uk
	----- Genome Center
COMMENT	Center: Sanger Centre
	Center code: SC
	Web site: http://www.sanger.ac.uk
	Contact: humquery@sanger.ac.uk
	----- Project Information
	Center project name: d0889P23
	----- Summary Statistics
	Assembly program: XGAP4; version 4.5
	Sequencing vector: M13; M7815; 12% of reads
	Sequencing vector: plasmid; L08752; 87% of reads
	Chemistry: Dye-terminator ABI; 14% of reads
	Chemistry: Dye-terminator Big Dye; 60% of reads
	Chemistry: Dye-Primer-amersham; 24% of reads
	Consensus quality: 131458 bases at least Q40
	Consensus quality: 131709 bases at least Q30
	Consensus quality: 131965 bases at least Q20
	Insert size: 133292; sum-of-contigs
	Insert size: 133129; 14.2% error; agarose-fp
	Quality coverage: 11.37x in Q20 bases; sum-of-contigs Quality
	coverage: 11.65x in Q20 bases; agarose-fp
	-----
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 4 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	-----
	* 1 3331: contig of 3331 bp in length
*	3332 3431: gap of 100 bp
*	3432 90998: contig of 87567 bp in length
*	90999 91098: gap of 100 bp
*	91099 130106: contig of 39008 bp in length
*	130107 130206: gap of 100 bp

\* 130207 132592: contig of 2386 bp in length.  
Location/Qualifiers  
1..132592  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.21-36.33"  
/clone="RP5-889P23"  
/clone\_1ib="RPC1-5"  
1..3331  
/note="assembly\_fragment:00799"  
3432..90998  
/note="assembly\_fragment:02114"  
91099..130106  
/note="assembly\_fragment:03033.0"  
130207..132592  
/note="assembly\_fragment:03380"  
ORIGIN  
Query Match 3.9%; Score 60; DB 2; Length 132592;  
Best Local Similarity 65.9%; Pred. No. 0.072;  
Matches 87; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 533 GCACGAGTGTGCGTGTGCGGAGCGGAGCTCATCTGTGACGCGCTGCCCTCG 592  
33964 GGAAGTTCTCCCGGCTGTGCAAGGCGGGCGAGCTCTGTGTGAGACGCTCCCTC 34023  
DB 593 GGCCTTCACCTGGCTGCTGCTCCCTCCGCTCCGAGATCCCAAGTGGAGCTGGAG 652  
34024 CTCCTACCACTGATGCTCAACCGCGCTGCCGAGATCCCAACGATGAATGCT 34083  
QY 653 GTGCTCCAGCTG 664  
DB 34084 CTGCCCGCGCTG 34095  
RESULT 68  
HS120G22 166518 bp DNA linear PRI 21-JUL-2000  
LOCUS Human DNA sequence from clone RP1-120G22 on chromosome  
DEFINITION  
1p36.21-36.33, complete sequence.  
ACCESSION  
AL031847  
VERSION  
AL031847.17 GI:9369286  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
Submitted (20-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requesters: clonerequest@sanger.ac.uk  
On Jul 22, 2000 this sequence version replaced gi:9368784.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL, Sw: SWISSPROT, Tr: TRMBL, Wp: WORMPEP, Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP1-120G22 is from the library RP1-1 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RP1-120G22 it may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true right end of clone RP1-120G22 is at 166518 in this  
sequence. The true left end of clone RP5-889P23 is at 90508 in this  
sequence. The true right end of clone RP1-20208 is at 100 in this  
sequence.  
FEATURES  
source  
1..166518  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="p36.21-36.33"  
/clone="RP1-120G22"  
/clone\_1ib="RPC1-1"  
8814..9131  
/note="match: GSS: Em:A0716455"  
8816..9229  
/note="match: GSS: Em:A0239786"  
10012..10541  
/note="match: GSS: Em:A0678815"  
18389..18823  
/note="match: GSS: Em:A0270793"  
25588..26045  
/note="match: GSS: Em:A0214655"  
25694..25967  
/note="match: GSS: Em:A0097494"  
complement(41087..41505)  
/note="match: GSS: Em:A0213363"  
complement(41123..41505)  
/note="match: GSS: Em:A0133445"  
45297..45591  
/note="match: GSS: Em:AP046498"  
49180..49534  
/note="Single clone region. short insert library only"  
complement(57981..58764)  
/note="match: GSS: Em:AP158011"  
complement(60582..61043)  
/note="match: GSS: Em:A0695221"  
64797..64841  
/note="Single clone region. short insert library only"  
68250..68620  
/note="match: STS: Em:G16281"  
70460..70622  
/note="match: STS: Em:G21080"  
complement(74594..74895)  
/note="match: GSS: Em:A0097224"  
complement(79470..79747)  
/note="match: STS: Em:AP191963"  
complement(79556..79740)  
/note="match: GSS: Em:A0541496"  
complement(79935..80319)  
/note="match: STS: Em:HS4052WJ1"  
complement(85942..86372)  
/note="match: STS: Em:G22706"  
complement(86963..87322)  
/note="match: GSS: Em:A0191756"  
87339..87808  
/note="match: GSS: Em:A0698012"  
88780..89285  
/note="match: GSS: Em:A012434"  
94030..94434  
/note="match: GSS: Em:B89150"  
94030..94405

```

misc_feature      /note="match: GSS: Em:AQ008338"
                    94054..94464
                    /note="match: GSS: Em:AQ010380"
                    join(112106..112211,116809..116933,123047..123132)
misc_feature      /note="match: GSS: Em:AQ634365"
                    complement(join(112121..112315,116809..116936))
misc_feature      /note="match: GSS: Em:AQ202845"
                    112126..112224
                    /note="match: GSS: Em:AF046365"
                    join(112145..112211,116809..116936)
                    /note="match: GSS: Em:AQ634370"
                    complement(join(112150..112211,116809..116933),
                    123047..123172))
misc_feature      /note="match: GSS: Em:AQ489599"
                    complement(113100..113712)
                    /note="match: GSS: Em:AQ477048"
                    complement(113175..113715)
                    /note="match: GSS: Em:AQ476615"
                    113710..114090
                    /note="match: GSS: Em:AQ434236"
                    join(116809..116933,123047..123119)
                    /note="match: GSS: Em:AQ732842"
                    116809..116941
                    /note="match: GSS: Em:AQ357462"
                    116809..116936
                    /note="match: GSS: Em:AQ047297 Em:AQ351276"
                    join(116840..116933,123047..123191)
                    /note="match: GSS: Em:G13324 Em:G13474"
                    121507..121762
                    /note="match: GSS: Em:AQ482227"
                    complement(123035..123248)
                    /note="match: GSS: Em:AQ779080"
                    123142..123248
                    /note="match: GSS: Em:AQ634365"
                    123192..123356
                    /note="match: GSS: Em:G11177"
                    124453..124584
                    /note="match: GSS: Em:AQ002260"
                    complement(124491..124832)
                    /note="match: GSS: Em:G06939"
                    129943..130018
                    /note="single clone region"
                    144105..144499
                    /note="match: GSS: Em:AQ219599"
                    144105..144483
                    /note="match: GSS: Em:AQ219566"
                    complement(148672..149104)
                    /note="match: GSS: Em:AQ672446"
                    152031..152510
                    /note="match: GSS: Em:AQ234353"

```

```

ORIGIN
Query Match      3.9%; Score 60; DB 9; Length 166518;
Best Local Similarity 65.9%; Pred. No. 0.07;
Matches 87; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 533 GACAGAGTGTGCGGTGTGTGCGAGCGGCGGAGCTCATCTGTGTGACGGCTGCCCTCG 592
    |||
    |||
    |||
Db 160874 GAGATTCTGCGCGGTGTGCAAGGACGGGGCGAGTCTGTGTGCGACGGCTGCCCTC 160933
    |||
    |||
    |||
QY 593 GGCCTTCCACCTGCGCTGTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGACCTCGAG 652
    |||
    |||
    |||
Db 160934 CTCCTACCACTGCTGCTCAACCGCGCTGCCCGAGATCCCAACGGTGAATGGCT 160993
    |||
    |||
    |||
QY 653 GTGCTCCAGCTG 664
    |||
    |||
    |||
Db 160994 CTGCCCGCGCTG 161005

```

```

RESULT 69
LOCUS AR383186 1053 bp mRNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6610823.

```

```

ACCESSION AR383186
VERSION AR383186.1 GI:40092750
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1053)
AUTHORS Targoff,I.N. and Ge,Q.
TITLE Antigen associated with polymyositis and with dermatomyositis
JOURNAL Patent: US 6610823-A 1 26-AUG-2003;
FEATURES
source 1..1053
    /organism="unknown"
    /mol_type="mRNA"

```

```

ORIGIN
Query Match      3.9%; Score 59.8; DB 6; Length 1053;
Best Local Similarity 58.9%; Pred. No. 0.17; 72; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGACAGAGTGTGCGGTGTGTGCGGACGGGCGGAGCTCATCTGTGTGACGGCTGCCCTC 591
    |||
    |||
    |||
Db 248 AGATTACTGTGAGGTGTGCGACGAGGAGGGAATTATCTGTGTGACACTGCCCTC 307
    |||
    |||
    |||
QY 592 GGCCTTCCACCTGCGCTGTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGACTTGA 651
    |||
    |||
    |||
Db 308 GTGCTTCCACCTGCGCTGTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGACTTGA 367
    |||
    |||
    |||
QY 652 GGTGCTCCAGCTGCTGCTGCGGCAACAGTCCAGAGAGGTGACGCCCGCGGAGAGA 706
    |||
    |||
    |||
Db 368 GCTGCCCTCACTGTGTGAGAGAGGGGCTCAAGTGGAGCCCAAGAGAGAGAGA 422

```

```

RESULT 70
HSU08379 1591 bp mRNA linear PRI 26-OCT-1995
LOCUS HSU08379
DEFINITION Human M1-2 autoantigen 240 kDa protein mRNA, partial cds.
ACCESSION U08379
VERSION U08379.1 GI:761717
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1591)
AUTHORS Mammla; Butcheria; Primates; Catarhni; Homindae; Homo.
TITLES Ge,Q., Nilasena,D.S., O'Brien,C.A., Frank,M.B. and Targoff,I.N.
    Molecular analysis of a major antigenic region of the 240-kD
    protein of M1-2 autoantigen
JOURNAL U. Clin. Invest. 96 (4), 1730-1737 (1995)
MEDLINE 96013633
PUBMED 7560064
REFERENCE 2 (bases 1 to 1591)
AUTHORS Frank,M.B.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1994) M. Bart Frank, Oklahoma Medical Research
    Foundation, Arthritis and Immunology Program, 825 NE 13th Street,
    Oklahoma City, OK 73104, USA
FEATURES
source 1..1591
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="C4"
    /cell_type="thymocyte"
    /tissue_type="thymus"
    /clone_lib="thymocyte and 5' stretch thymocyte libraries
    (Clontech)"
    <1..>1591
    /note="Northern blot with this cDNA shows hybridization of
    a single band in 7.5 kb region"
    /evidence="experimental"
    <1..>1591
    /note="The partial cDNA encodes an antigenic region of a

```



	nuclear protein to which autoantibodies are produced by 15-20% of dermatomyositis patients"
	/codon_start=2
	/product="M1-2 autoantigen 240 kDa protein"
	/protein_id="AAC50228.1"
	/db_xref="gi:761718"
	/translation="PHFOKOKOYVEOKSATLLTWGLDEVHVFSESDYHTLTNYAKA FSQMRPLIAKNPKIPMSKMKTILGAWREFSANNPKSGAAVAAAAVAALAE OSVAUSATPIAPSGPALPPPAADLOPPIRATRTBEKGSHGRASPRVPVG RKLRGKKMAFLKIHLGLGSKRKGGSSDGPPELFAEBESLDJGSVHASAGRDGVF RTKLRLPERPERRKKRVLGCPAVAGEBESVDVETDHODYCEVCQQGEILLCDTPCA YHALCDELBDLPRAPEGKWSCPCHKEAQWQMAKEBEEBEGEGBEKEDMDMEYCV CRVKDGSELICDCACISYSIHCLNLPDPPIPGSEWLCPCTCPVLKGRVQKLIHRM WGSEPVPAPAPOAADGNVDVPPRPLOGRSREPFVKVGVSYNGCSAKELOLEIPEH LMVRNTYORKNDMBEPPLIDYGSGBEDGSKDRKRVKXOPHYALEMEKYYRIRIKEMMT VHRLINHSVDKKNGNYHYLVKR"
ORIGIN	
Query Match	3.9%; Score 59.8; DB 9; Length 1591;
Best Local Similarity	58.9%; Pred. No. 0.16;
Matches 103; Conservative	0; Mismatches 72; Indels 0; Gaps 0;
OY	532 AGAGCGATGTGCCGTGTGTGCGGAGCGGC GGGAAGACTATCTGCTGTACAAGCTGCCCTC 591
DB	765 AGGATTACTGTGAGGTGTGCCAGCGAGGGTGGGAAATTATTCGTGTACACCTGCCCTC 824
OY	592 GGGCCCTTCCAGCTGGCCTGCTGCTCCCTCCGCTCCGGAGAATCCCCAGTGAGGACCTTGA 651
DB	825 GTGCTTACCACCTCTCTGTGCTTTAGTCTTAGAGCTTGACCGGGCTTCAGAGGGCAAATGA 884
OY	652 GTGTCTCAGCTGCTTGCAGAGCACAGTTCAGAGAGTGACAGCCCGGSCAGAGA 706
DB	885 GCTGCCCTCACCTGTGAGGAAGAGGGGGTCTCAGTGAGGAGCCAGAGGAGAAAGA 939
RESULT 71	
LOCUS	CQ491372 6311 bp DNA linear PAT 30-JAN-2004
CQ491372	
DEFINITION	Sequence 23239 from Patent WO0160860.
ACCESSION	CQ491372
VERSION	CQ491372.1 GI:41456991
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 Schlegel R., Endege W.O. and Monahan J.E. Genes differentially expressed in human prostate cancer and their use
AUTHORS	Patent: WO 0160860-A 23239 23-AUG-2001;
TITLE	Millennium Predictive Medicine, Inc. (US) Location/Qualifiers
JOURNAL	1..6331
FEATURES	/organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"
SOURCE	
ORIGIN	
Query Match	3.9%; Score 59.8; DB 6; Length 6331;
Best Local Similarity	58.9%; Pred. No. 0.13;
Matches 103; Conservative	0; Mismatches 72; Indels 0; Gaps 0;
OY	532 AGAGCGATGTGCCGTGTGTGCGGAGCGGC GGGAAGACTATCTGCTGTACAAGCTGCCCTC 591
DB	1346 AGGATTACTGTGAGGTGTGCCAGCGAGGGTGGGAAATTATTCGTGTACACCTGCCCTC 1405
OY	592 GGGCCCTTCCAGCTGGCCTGCTGCTCCCTCCGCTCCGGAGAATCCCCAGTGAGGACCTTGA 651
DB	1406 GTGCTTACCACCTCTCTGTGCTTTAGTCTTAGAGCTTGACCGGGCTTCAGAGGGCAAATGA 1465
OY	652 GTGTCTCAGCTGCTTGCAGAGCACAGTTCAGAGAGTGACAGCCCGGSCAGAGA 706

Db	1466	GCTCCCTCACTGTGAGAGAGGAGGGGTCCAGTGTGGAGGCCAAGAGAGAAAGA	1520
RESULT 72			
LOCUS	AX924075	6331 bp	linear
DEFINITION	Sequence 29 from Patent WO03080105.		PAT 18-DEC-2003
ACCESSION	AX924075		
VERSION	AX924075.1	GI:40217079	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1		
FEATURES	Al-Mahmood,S., Colin,S. and Schneider,C.		
source	Angiogenesisis regulator genes, pharmaceutical preparations		
	containing same and uses thereof		
	Patent: WO 03080105-A 29 02-OCT-2003;		
	Gene Signal (FR)		
	Location/Qualifiers		
	1..6331		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	3.9%; Score 59.8; DB 6; Length 6331;		
Best Local Similarity	58.9%; Pred. No. 0.13; 72; Indels 0; Gaps 0;		
Matches	103; Conservative 0; Mismatches		
Qy	532	AGACGAGGTGCGCGTGTGCGGAGCGCGGGAGCTCATCTGCTGACGGCTCCCTC	591
Db	1346	AGGATTACTGTGAGGTGTGCCAGCAGGCTGGGGAAATTAATCTGTGTGACACCTGCCCC	1405
Qy	592	GGGCTTCCACCTGCGCTGCTGCTCCCTCCCTCCGCTCCGGAGATCCCGAGTGGACCTGGA	651
Db	1406	GTGCTTACCACTTGTCTGCTGCTTATCTTGAGCTTGACCGGCTTCCAGAGGCAAAATGA	14655
Qy	652	GGTCTCTCAGCTGCTGCTGAGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA	706
Db	1466	GCTCCCTCACTGTGAGAGAGGAGGGGTCCAGTGTGGAGGCCAAGAGAGAAAGA	1520
RESULT 73			
LOCUS	AX924076	6331 bp	linear
DEFINITION	Sequence 30 from Patent WO03080105.		PAT 18-DEC-2003
ACCESSION	AX924076		
VERSION	AX924076.1	GI:40217080	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1		
FEATURES	Al-Mahmood,S., Colin,S. and Schneider,C.		
source	Angiogenesisis regulator genes, pharmaceutical preparations		
	containing same and uses thereof		
	Patent: WO 03080105-A 30 02-OCT-2003;		
	Gene Signal (FR)		
	Location/Qualifiers		
	1..6331		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	3.9%; Score 59.8; DB 6; Length 6331;		
Best Local Similarity	58.9%; Pred. No. 0.13; 72; Indels 0; Gaps 0;		
Matches	103; Conservative 0; Mismatches		
Qy	532	AGACGAGGTGCGCGTGTGCGGAGCGCGGGAGCTCATCTGCTGACGGCTCCCTC	591





Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idelbird, D., Jackson, A., Jacob, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Nandasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelamen, O., Okunolu, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plodner, F., Poindexter, A., Popovic, D., Prym, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. U., Sanders, M., Savary, G., Scherer, S., Scott, G., Shattman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valse, R., Vere, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Wortley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, D., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

**TITLE**  
Direct Submission  
Unpublished

**JOURNAL**  
2 (bases 1 to 168479)

**REFERENCE**  
Rat Genome Sequencing Consortium.

**AUTHORS**  
Submitted (21-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**JOURNAL**  
3 (bases 1 to 168479)

**REFERENCE**  
Rat Genome Sequencing Consortium.

**AUTHORS**  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**JOURNAL**  
On Nov 15, 2002 this sequence version replaced gi:24181980. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

**----- Genome Center**  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

**----- Project Information**  
Center project name: KCMV  
Center clone name: CH230-368023

**----- Summary Statistics**  
Assembly program: Phrap, version 0.990329  
Consensus quality: 163230 bases at least Q40  
Consensus quality: 165491 bases at least Q30  
Consensus quality: 166611 bases at least Q20  
Estimated insert size: 169544; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

**NOTE:** Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_drat\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_drat_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced by the finished sequence as soon as it is available and \* the accession number will be preserved.

**FEATURES**  
source  
Location/Qualifiers  
1..168479  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-368023"  
/note="CH2776..3643"  
/complement(12776..3643)  
/note="clone\_boundary"  
clone\_end:T7  
site:  
end\_sequence:B2145823"  
/complement(166515..167327)  
/note="clone\_boundary"  
clone\_end:Sp6  
site:  
end\_sequence:B2145824"

**ORIGIN**  
Query Match 3.9%; Score 59.8; DB 2; Length 168479;  
Best Local Similarity 55.6%; Pred. No. 0.076;  
Matches 115; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

**misc\_feature**  
complement(166515..167327)  
/note="clone\_boundary"  
clone\_end:Sp6  
site:  
end\_sequence:B2145824"

**misc\_feature**  
complement(12776..3643)  
/note="clone\_boundary"  
clone\_end:T7  
site:  
end\_sequence:B2145823"

**misc\_feature**  
complement(166515..167327)  
/note="clone\_boundary"  
clone\_end:Sp6  
site:  
end\_sequence:B2145824"

**RESULT 78**  
AC119782/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-134W24, WORKING DRAFT SEQUENCE.  
AC119782  
AC119782.5 GI:23670853  
VERSION  
HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 221787)  
Muzny, D., Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behnmed, F., Bielawski, K., Blais, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garsa, W., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kow, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervyls, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemele, O., Okunolu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzio, M., Quiroz, J., Rachlin, E., Reeves, K., Register, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rutz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Silter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

**TITLE**  
Unpublished  
2 (bases 1 to 221787)  
**REFERENCE**  
Worley, K.C.  
**AUTHORS**  
Direct Submission  
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 221787)  
**JOURNAL**  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Oct 10, 2002 this sequence version replaced gi:21903167.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GWIP  
Center name: CH230-134M24  
----- Summary Statistics  
Assembly program: PIRAP; version 0.990329  
Consensus quality: 207183 bases at least Q40  
Consensus quality: 208452 bases at least Q30  
Consensus quality: 209456 bases at least Q20  
Estimated insert size: 215794; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 221787: contig1 of 221787 bp in length.  
Location/Qualifiers

source  
1. 221787  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-134M24"  
1. 1979  
/note="wgs contig"  
complement(220394..221263)  
/note="clone\_boundary  
clone\_end:77  
site:ECORI  
end\_sequence:BH349190"

## ORIGIN

Query Match 3.9%; Score 59.8; DB 2; Length 221787;  
Best Local Similarity 55.6%; Pred. No. 0.073; Indels 0; Gaps 0;  
Matches 115; Conservative 0; Mismatches 92;

QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGAGCTCATCTGTCGACGCTGCGCTCG 592  
DB 66840 GGAATTTTCGCCGCGTGTGCAAGACGTCGTCGTGTGATGCTTGTCCCTC 66781  
QY 593 GGCCTTCCACCTGCGCTGCTTCTCCCTTCGCTCCGAGATCCCACTGAGAC 652  
DB 66780 CTCCTACCCCTTGCACCTGCTCAACCCGCGCTGCGGAGATCCGAAAGTGAATGCT 66721  
QY 653 GTGCTCCAGCTCCCTGCAAGCAAGTCCAGAGGTCGAGCCCGGAGAGAGCCCG 712  
DB 66720 CTGCGCGCGCTTACAGTGAAGTTCATCTTCTGCGCCGCGCCCGCCCGCCG 66661  
QY 713 GCCCGAGAGCCACCCGTCGAGACCC 739  
DB 66660 CCG 66624

RESULT 79  
AC097926/c 246566 bp DNA linear HTG 10-MAY-2003  
LOCUS AC097926/c  
DEFINITION Rattus norvegicus clone CH230-44U16, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
ACCESSION AC097926  
VERSION AC097926.6 GI:30521120  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 246566)









Takeuchi, K., Arita, M., Imose, N., Muraahino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Komiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujimori, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiroto, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kodatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunuma, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, K., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okunura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuo, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)

JOURNAL  
PUBMED  
14702039  
TITLE  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 6434)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Chisara, Tel:81-438-52-3975, Fax:81-438-52-3986  
(E-mail:genominfo@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.

FEATURES  
source  
Location/Qualifiers  
1..6434  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UTERUS302536"  
/issue\_type="uterus"  
/clone\_lib="UTERUS3"  
/note="Cloning vector: pME18SFL3"

ORIGIN  
Query Match  
Best Local Similarity 3.8%; Score 58.8; DB 9; Length 6434;  
Matches 96; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

532 AGGACGAGTGTCCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 591  
1211 AGGACTATTCGAGGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 1270  
592 GGGCTTCACCTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 651  
1271 GTGCTTACCACTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 1330  
652 GTGCTTCACCTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 689  
1331 GTGCCACACTGCGGAGCGGAGCGGAGCTCATCTGTGACGCGCTCC 1368

RESULT 83  
LOCUS  
AL611985  
DEFINITION  
Mouse DNA sequence from clone RP23-421E12 on chromosome 4, complete sequence.  
ACCESSION  
AL611985  
VERSION  
AL611985.22 GI:23304628  
KEYWORDS  
HTG.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
Bates, K.  
Submitted (19-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Sep 23, 2002 this sequence version replaced gi:22797927.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
RP23-421E12 is from the RP23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6.

FEATURES  
source  
Location/Qualifiers  
1..219205  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-421E12"  
/clone\_lib="RP23-421E12"

ORIGIN  
Query Match  
Best Local Similarity 3.8%; Score 58.4; DB 10; Length 219205;  
Matches 86; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

533 GGAAGTGTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 592  
101700 GGAAGTGTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 101759  
593 GGGCTTCACCTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 652  
101760 CTCTTACCACTGCGCTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCC 101819  
653 GTGCTTCACCTG 664  
101820 CTGCCGCGCTG 101831

```

RESULT 84
LOCUS      D86113              63 bp      DNA              linear      PRI 04-JUL-1997
DEFINITION Homo sapiens gene for HC21EXC33, exon.
ACCESSION  D86113
VERSION     D86113.1 GI:2244722
KEYWORDS   HC21EXC33.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Mammalia; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 63)
AUTHORS     Kudoh,J., Nagamine,K., Asakawa,S., Abe,I., Kawaaki,K., Maeda,H.,
            Tsujimoto,S., Minoshima,S., Ito,F. and Shimizu,N.
TITLE       Localization of 16 exons to a 450-kb region involved in the
            autoimmune polyglandular disease type I (APECED) on human
            chromosome 21q22.3
JOURNAL     DNA Res. 4 (1), 45-52 (1997)
MEDLINE     97333005
PUBMED      9179495
REFERENCE   2 (bases 1 to 63)
AUTHORS     Nagamine,K., Kudoh,J., Asakawa,S., Abe,I., Maeda,H., Tsujimoto,S.,
            Minoshima,S., Ito,F. and Shimizu,N.
TITLE       Localization of 22 Exons to a 450-kb Region Involved in the
            Autoimmune Polyglandular Disease Type I (APECED) on Chromosome
            21q22.3
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 63)
AUTHORS     Shimizu,N.
TITLE       Direct Submision
JOURNAL     Submitted (19-JUN-1996) Nobuyoshi Shimizu, Keio University School
            of Medicine, Department of Molecular Biology, 35 Shinanomachi,
            Shinjuku-Ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp,
            Tel:03-3351-2370, Fax:03-3351-2370)
            Location/Qualifiers
FEATURES
    source
        1..63
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /cell_line="GM130B"
            /cell_type="B-Lymphoblastoid"
            /clone_11b="chromosome 21-specific cosmid library"
    gene
        1..63
            /gene="HC21EXC33"
    exon
        1..63
            /gene="HC21EXC33"
    ORIGIN
        Query Match      3.8%; Score 58.2; DB 9; Length 63;
        Best Local Similarity 95.2%; Pred. No. 0.56;
        Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
    Oy      1149 GATGACATCGCCAGCAGCGCCGCTGCA CAGGAGTGAAGTCCCTTGAGC 1208
            |||
            1 GATTGACACGGCAGTCAAGGCGCTCTGCACAGGAGTGAAGTCCCTTGAGC 60
    Db      1209 GAG 1211
            |||
            61 GAG 63

```

```

SOURCE
ORGANISM    Mus musculus (house mouse)
MUS musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4343)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,K.H., Scheaffer,C.F., Bhat,N.K.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Uebelin,T.B., Toshiyuki,S.,
            Carinici,P., Prange,C., Raha,S.S., Loquillano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalón,D.K., Wuzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Hellon,E., Kettelman,M., Madan,A., Rodriguez,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schultz,J., Myers,R.M.,
            Butlerfield,Y.S., Krzywinski,M.I., Skalek,U., Smalins,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE       Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED      12477932
REFERENCE   2 (bases 1 to 4343)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-NOV-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Dr. Jim Ivin, University of Iowa
            CDNA Library Preparation: M. Benito Soares, University of Iowa
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www.shgc.stanford.edu
            Contact: (Dickson, Mark) mc@paxll.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clome distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAX Plate: 125 Row: f Column: 19
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES
    source
        1..4343
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:6406263"
            /tissue_type="Brain, enriched mouse brain 12.5dp"
            /clone_11b="NIH BMAP_F00"
            /lab_host="DH10B"
            /note="Vector: pYX-ASC"
    ORIGIN
        Query Match      3.8%; Score 58.2; DB 10; Length 4343;
        Best Local Similarity 58.3%; Pred. No. 0.28;
        Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
    Oy      532 AGGACGAGTGGCCGTGTGTGGGACGCGCGGAGAGCTATCTGCTGACGCGCTCCCTC 591
            |||
            512 AGGATTACTGTGAGGTGTGCAGCAGGCGGCGGGAATTATCTGTGCGACACTGCCCC 571

```

QY 592 GGGCCCTCCAGCTGCGCTGCTCCCTCCGTCGGAGATCCCAAGTGGAGCACTGGA 651  
|||||  
Db 572 GTGGCTACACCTCGCTGCTGCTGAGACCCGAGCTTGAACCGGCTGAGAGGCAAGTGA 631  
|||||  
QY 652 GGTGCTCCAGCTGCTGCGAGCAAGTCCAGAGAGTGCAGCCCGGCGAGAGCA 706  
|||||  
Db 632 GCTGCCCTCAGCTGTGAGAGAGAGGAGGTGAGGAGCCCAAGAGAGAGAGCA 686  
|||||

RESULT 86  
AC133595/LOCUS  
DEFINITION Homo sapiens chromosome 17 clone RP11-242A8 map 17, 3 unordered  
pieces.  
AC133595 116984 bp DNA 1linear HTG 27-MAR-2003  
AC133595.14 GI:27452934  
HTG: HTGS\_PHASE1; HTGS\_FULFLOP; HTGS\_CANCELLED.  
VERSION  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Birren, B., Nussbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 17, clone RP11-242A8  
JOURNAL  
REFERENCE  
AUTHORS Unpublished  
2 (bases 1 to 116984)  
Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarat, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand, Piere, N., Hago, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,  
McCarthy, M., Meldrum, J., Meneu, L., Mithova, T., Mlenga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Piere, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,  
Teste, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.  
DIRECT SUBMISSION  
Submitted (14-SEP-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 116984)  
Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
Camarat, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand, Piere, N., Hago, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
Mathews, C., McCarthy, M., Meldrum, J., Meneu, L., Mithova, T.,  
McCarthy, M., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Piere, N., Raymond, C., Retta, R.,  
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N.,  
Stojanovic, N., Talamas, J., Teste, S., Theodore, J., Topham, K.,  
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
DIRECT SUBMISSION  
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jan 1, 2003 this sequence version replaced gi:27369450.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L28162  
Center clone name: 242\_A\_8  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 40901: contig of 40901 bp in length  
\* 40902 41001: gap of 100 bp  
\* 41002 78861: contig of 37860 bp in length  
\* 78862 78961: gap of 100 bp  
\* 78962 116984: contig of 38023 bp in length.  
FEATURES  
source  
1..116984  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-242A8"  
/clone\_11b="RPCT-11 Human Male BAC"  
ORIGIN  
Query Match 3.8%; Score 58.2; DB 2; Length 116984;  
Best Local Similarity 59.3%; Pred. No. 0.17; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 68;  
QY 532 AGCAGAGTGTGCGCTGCTGCGAGACGGGAGACTCATCTGCTGACGCGCTGCTC 591  
|||||  
Db 31093 AGCATTACTGTGAGGTGCGAGACGGGAGGGAATTTCTGTGTCACACCTGCCCC 31034  
|||||  
QY 592 GGGCTTCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCACTGGA 651  
|||||  
Db 31033 GTGGCTACACCTCGCTGCTGCTGAGACCCGAGCTTGAACCGGCTGAGAGGCAATGGA 30974  
|||||  
QY 652 GGTGCTCCAGCTGCTGCGAGCAAGTCCAGAGAGTGCAGCCCGGCGAGAGCA 706  
|||||  
Db 30973 GCTGCCCTCAGCTGTGAGAGAGAGGAGGTGAGGAGCCCAAGAGAGAGCA 686  
|||||

RESULT 87  
AC104581  
LOCUS  
DEFINITION Homo sapiens chromosome 17, clone RP11-1099W24, complete sequence.  
AC104581 187718 bp DNA 1linear PRI 13-MAY-2003  
AC104581.22 GI:30581670  
VERSION  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Birren, B., Nussbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 17, clone RP11-1099W24  
JOURNAL  
REFERENCE  
AUTHORS Unpublished  
2 (bases 1 to 187718)  
Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarat, J., Campopiano, A., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collymore, A., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPeeters, R., Meldrum, D., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, W., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trillio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (13-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 187718)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Batra, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 187718)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Batra, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Submitted (13-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 13, 2003 this sequence version replaced gi:29150453.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)  
----- Project Information  
Center project name: L22268  
Center clone name: 1099\_M\_24  
-----

FEATURES	source
location/Qualifiers	1..187718
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="17"	
/map="17"	
/clone="RP11-1099M24"	
/clone_11b="RP11 Human Male BAC"	
789..1020	
/rpt_family="MIR"	
1539..1582	
/rpt_family="CGA n"	
1675..1676	
/note="unresolved homopolymeric run: restriction enzyme digest fingerprint indicates about bp of poly C repeat missing here."	
4209..4468	
/rpt_family="MIR"	
4732..4902	
/rpt_family="CAGAGA n"	
complement(7246..7250)	
/note="30 qual SNGL region"	
10077..10135	
/rpt_family="CCA n"	
10625..10654	
/rpt_family="CAG n"	
10880..10977	
/rpt_family="C-rich"	
11178..11217	
/rpt_family="CAG n"	
11775..11817	
/rpt_family="CCA n"	
11836..11918	
/rpt_family="C-rich"	
16622..16642	
/rpt_family="GC-rich"	
16905..16934	
/rpt_family="AT-rich"	
complement(16991..17008)	
/note="30 qual SNGL region"	
17644..17674	
/rpt_family="AT-rich"	
18005..18020	
/rpt_family="AT-rich"	
complement(23383..23647)	
/rpt_family="AluJo"	
complement(23661..23811)	
/rpt_family="MIR"	
24851..25513	
/rpt_family="12"	
25693..26008	
/rpt_family="AluS4"	
complement(27601..27833)	
/rpt_family="AluJo"	
complement(27834..27858)	
/rpt_family="Tigget4(Zombi)"	
27859..28148	
/rpt_family="AluSx"	
complement(28149..28277)	
/rpt_family="Tigget4(Zombi)"	
complement(28432..28554)	
/rpt_family="MIR"	
complement(28635..28988)	
/rpt_family="MLTIG1"	
complement(29134..29251)	
repeat_region	



```

Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GMD
Center clone name: CH230-150P11
-----
Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 198152 bases at least Q40
Consensus quality: 199686 bases at least Q30
Consensus quality: 200691 bases at least Q20
Estimated insert size: 204955; sum-of-coverage
Quality coverage: 7x in Q20 bases; sum-of-coverage
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 26928: contig of 26928 bp in length
* 26929 27028: gap of unknown length
* 27029 231727: contig of 204699 bp in length
* 231728 231827: gap of unknown length
* 231828 232984: contig of 1157 bp in length.
Location/Qualifiers
source
1. 232984
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-150P11"
misc_feature
1. 1443
/note="wgs end-extension
clone_end:T7"
misc_feature
456. 843
/note="clone boundary
clone_end:T7
site:BCORI
end_sequence=BH333066"
ORIGIN
Query Match
Best Local Similarity 64.0%; Pred. No. 0.19; Length 232984;
Matches 87; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 529 ATGAGACGAGTGTGCGGTGTCGGGACGGCGGAGGCTCATCTGTGTGACGGCTGCC 588
Db 72989 ATATGGAATTCTGTGCAATGTGCAAGATGCGGGAACTACTATCTGTGACACTGCC 73048
QY 589 CTCGGGCGCTTCACCTGCGCTGCTCCCTCCGCTCCGGAGAGATCCCGACGAGCACT 648
Db 73049 CTTCTTCTTACCAATATTCACCTGTGTGAACCCCGCTGCGAGAGATCCAAATGGCAAT 73108
QY 649 GGAGGTGTCTCAGCTG 664
Db 73109 GGCTCTGTCCCGCTG 73124
-----
RESULT 89
AC115415 261384 bp DNA linear HTG 21-SEP-2002
AC115415 Rattus norvegicus clone CH230-164A18, *** SEQUENCING IN PROGRESS
DEFINITION
*** 2 unordered pieces.
ACCESSION
AC115415
VERSION
AC115415.3 GI:23266199
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 261384)
Muzny,D,Marie, Metzker,M, Lee, A, Brannon, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amun, A., Anguiano, D.,
Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceaar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De And, C., Dederich, D.,
Delgado, O., Denon, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idledit, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensbaw, L., Louised, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindarane, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Miklosavljovic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munda, M., Murphy, M., Natr, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaekeme, O., Okumu, G., Olarunagbon, A., Pal, S., Parke, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shwartsbeyn, A., Sison, I., Sitter, C. D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, J., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 261384)
Worley, K. C.
Direct Submission
Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 261384)
Baylor Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21737225.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/atl). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
-----
Genome Center

```

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----  
 Center project name: GONQ  
 Center clone name: CH230-164A18

----- Summary Statistics -----  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 215637 bases at least Q40  
 Consensus quality: 217908 bases at least Q30  
 Consensus quality: 219565 bases at least Q20  
 Estimated insert size: 232237; sum-of-coverage estimation  
 Quality coverage: 5x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length -----  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

FEATURES  
 source  
 1 253628: contig of 253628 bp in length  
 253629 253728: gap of unknown length  
 253729 261384: contig of 7656 bp in length.  
 Location/Qualifiers  
 1. 261384  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-164A18"  
 1. 1140  
 /note="wgs\_end\_extension"  
 clone\_end="T7"  
 1665. 1950  
 /note="clone\_boundary"  
 clone\_end="T7  
 site:ECORI  
 end\_sequence="BH359670"  
 142337. 242648  
 /note="clone\_boundary"  
 clone\_end="Sp6  
 site:ECORI  
 end\_sequence="BH359671"

ORIGIN  
 Query Match 3.7%; Score 57.6; DB 2; Length 261384;  
 Best Local Similarity 64.0%; Pred. No. 0.19;  
 Matches 87; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

529 ATGAGAGCAGATGTGCGTGTGCGGAGCGGAGGAGCTCATCTGCTGACGCGTGC 588  
 162798 AATATGAAATTTCTGTGACAGTCTGACAGAGGCGGGAATCTACTATGCTGTGACACTTGC 162857

589 CTCGGGCTTTCCACCTGAGCTGCTGCTCCGCTCCGCGAGATCCCACTGGAGACCT 648  
 162858 CTTCTTCTTACCAATATTCACATGCTGTAACCCCGCTGCAGAGATCCCAATATGGGAAAT 162917

649 GGAGGTGCTCCAGCTG 664  
 162918 GGCTCTGTCTCCGCTG 162933

RESULT 90  
 AC123167/ 278068 bp DNA linear HTG 13-NOV-2002  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-70G17, \*\*\* SEQUENCING IN PROGRESS  
 AC123167  
 AC123167.4 GI:24635638

KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus (Norway rat)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus

REFERENCE  
 AUTHORS  
 1 (bases 1 to 278068)  
 Muzny, D., Marie, M., Metcalf, M., Lee, A., Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsebrook, S., Amin, A., Angiano, D.,  
 Anyalelechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bernadine, F.,  
 Bismar, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M., Davis, C., Davy-Carroll, L., De Andrade, C., Dederich, D.,  
 Delgado, O., Benson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregorjse, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Hayak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,  
 Hollins, B., Howells, S., Hui, S., Hume, D., Idolbit, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,  
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshew, L., Louissegh, H., Lozano, R.D., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Manning, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwalekeme, O., Okwou, G., Olarunuga, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, F., Polindexter, A., Popovic, D., Prims, E., Pu, L.,  
 Puazo, M., Quiroz, J., Rachlin, R., Reeves, K., Regier, M., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rivers, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajda, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,  
 Vales, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Wright, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von  
 Niederhausen, A., Weise, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstock, G. and Gibbs, R.A.

Direct Submission  
 Unpublished  
 2 (bases 1 to 278068)  
 Worley, K.C.  
 Direct Submission  
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 278068)  
 Submitted (13-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas







DVQGAHAQPTLQTPSIVQFGHHQKLSHFQOQPOOQLPPPPPLPPPPPLPPPP  
OOPHPPLPSOHLASOHSPPPPACSONMDIMHHEFLEEMOKLELLOOQPSLO  
LSOTSPHOILOOTIVGOIVIVROPAPOVOSOGSEITLOATDEBPASOGSKPALPLDKN  
TAAALPQASGEPLSPVPVDSTIOSSNNVVRKHSITSTISFSTLEMELSSTRLE  
RUEPOIQSVNLTGAPQAVPSILSPKMSVSLTSVONQAMPSTLTSHLQTPSLV  
HSTFQSPMLISDSFGAMASLDHDAQPSILMSGTQAVPSLATPLOSIPVSDMO  
PTGSSSSSGRTSGSLCPRDGADPSLEMLACRVSPEMLSKLPLFTIGOKIGHMDEYS  
DLSLTVLRPLMTTMSSEFDSCHFTFPERKVRILPLASLTMYDKVP"

## ORIGIN

Query Match 3.7%; Score 56.6; DB 9; Length 6505;  
Best Local Similarity 55.3%; Pred. No. 0.54;  
Matches 110; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 526 AGAATGAGGACGAGTGTGCGGAGGAGCGGAGGAGCTGCTGTGAGCGGT 585  
DB 3352 AGAATGAGGACCTTCTGTGCTTCTCAATGCGAGAGTTACTGTGTGAGCGGT 3411  
QY 586 GCCCTGCGGCTTCCACTGCTGCTGCTGCTGCTGCTGCGGAGATCCCGAGTGGGA 645  
DB 3412 GCCCAAGAGTTCGACCTCTCTGCGATGTGCGACCTTGTGCTGCTGCGAGGGAG 3471  
QY 646 CCTGAGGTGCTTCACTGCTGCGAGCAACAGTCCAGAGTGTGCGCGCGGAGAG 705  
DB 3472 AGTGGGTGTGTACTTGTGCGCGAGCTGACCGAGCCGAGATGAGTACTGTGAGA 3531  
QY 706 AGCCCGCGCGCGCGAGAGCC 724  
DB 3532 ATGCGCTGCTATACGAGCC 3550

RESULT 94  
AF210315 5811 bp mRNA linear INV 31-MAY-2001  
LOCUS Drosophila melanogaster bonus (bon) mRNA, complete cds.  
DEFINITION AF210315  
ACCESSION AF210315.1 GI:6630999  
VERSION AF210315.1  
KEYWORDS Drosophila melanogaster (fruit fly)  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 5811)  
AUTHORS Beckstead, R., Ortiz, J.A., Sanchez, C., Prokopenko, S.N., Chambon, P.,  
Losson, R. and Bellen, H.J.  
TITLE Bonus, a Drosophila homolog of TIF1 proteins, interacts with  
nuclear receptors and can inhibit betaFTZ-F1-dependent  
transcription  
JOURNAL Mol. Cell 7 (4), 753-765 (2001)  
MEDLINE 21235528  
PUBMED 11336699  
REFERENCE 2 (bases 1 to 5811)  
AUTHORS Beckstead, R.B., Prokopenko, S.N. and Bellen, H.J.  
TITLE Direct Submision  
JOURNAL Submitted (01-DEC-1999) Molecular and Human Genetics/HMTI, Baylor  
College of Medicine, T634, Mail Stop BCM335, One Baylor Plaza,  
Houston, TX 77030, USA  
FEATURES  
source location/Qualifiers  
1..5811  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/chromosome="3"  
/map="92E8-14"  
1..5811  
/gene="bon"  
1..693  
/gene="bon"  
694..4095  
/gene="bon"  
/note="TIF1-like protein; BON"  
/codon\_start=1

/product="bonus"  
/protein\_id="AAF19646.1"  
/db\_xref="GI:6631000"  
/translation="MDMDLEQKRLPLIAGIKQEOULDAVPTDALPOMSPNTASGA  
PTTSSVSSSLSPDCSAERKSSNSAKKFLFKCVYCAQLLGSNDRKLLKEL  
HVACQSSSTKSELDRLPLILHPVCDNMQSEIVDNPGLIEQTADEGDEVLG  
LGLTEGQSSASAIQCSGSDGVAIVSCVDCSEYICDSGVQKHQIKTKHTIK  
PDEANNLQALGAACVLDKLMQCLOPQKLSLFCCTCDKLTCRDCOLSDHDKKXFA  
HEIATESQALSTLSEINVKRFLSSATKVLVDQOOLIHDKKOLKEITMAVKIT  
NTVNRGKOLIMRINEVCDSKLVVEKKETLOLSDMTDCTIDMONALEKSGPAI  
LSSKSLVRLHOKLCKQRAIDIPNEIPIRIQUNOVSDLOKVISQLITTYDKPRP  
PLPSPNGTPQHPERPSPPMAPRLPGLPQMPAGLSPGPPVNGPQGPPLYS  
NAAQQFNNLSMSRSPGDSGKVRFGMPVPVGMQRHQPVSSSTHQNMDISLGL  
LNNQASQPNAMAFNCPSPYFGCPQAPADHPQMPQMPHFMPGQGFSGCGAG  
GSPRANPMNSNARFOSQYORMANNAOAAAHGAAGAGGAGQIIPSGALORPOM  
SNPMONVAMONSLGFHSGOAGFNTGPOTSPOLGCGGMSLAKKHTIPQASQSMCS  
OQGLPLPPRANQTSNFKTSLKSPNTLTNSTPBLGGGAGHOGHNGSSAQLNMA  
ALGLPAVAILSNVTSTLPTPSPSTHNTDFTPEPIDKVDSDINDILATLAKDSN  
GVQVLPBRTKTTSPQVHSTDLSNTQVNNKNEQKDPNEDMCAVCLDGLMLCDK  
CPKVPHQNCI PAISLPDESBSMCLLCVNIKELTTEGSEKSSGSLALELILQ  
RICLEIYOVEGSLNFRBESPAPNTSYEIVSSPMSLVIRRLDPSSPNHKDIAGF  
VSDVRLIRSNLYLFOEDTKYSNAKYLENPREPOLAKMLPQECTKXQKXNTSNP  
ALLGVNATGSPSPLENGKSCGSASLSDSDACLPKAKRBSAHE"

misc\_feature  
931..1092  
/gene="bon"  
/note="Region: ring-finger domain"  
misc\_feature  
1225..1338  
/gene="bon"  
/note="Region: B-box 1"  
misc\_feature  
1411..1506  
/gene="bon"  
/note="Region: B-box 2"  
misc\_feature  
1510..1929  
/gene="bon"  
/note="Region: coiled-coil domain"  
misc\_feature  
3385..3528  
/gene="bon"  
/note="Region: PHD-finger domain"  
misc\_feature  
3571..3909  
/gene="bon"  
/note="Region: bromodomain"  
3'UTR  
4096..5811  
/gene="bon"  
4543..4548  
/gene="bon"

ORIGIN  
polyA\_signal

Query Match 3.7%; Score 56.4; DB 3; Length 5811;  
Best Local Similarity 59.3%; Pred. No. 0.6;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 528 AATGAGGACGAGTGTGCGGAGGAGGAGGAGCTGCTGTGAGCGGTGC 587  
DB 3379 AATGAGGAGTGTGCGGCTGTGTGATGAGAGGAGCTATGTGCGACAAGTGT 3438  
QY 588 CCTGCGGCTTCCACTGCTGCTGCTGCTGCTGCTGCGGAGATCCCGAGTGGACC 647  
DB 3439 CCCAAGTTTCCATCAGAAAGTGTCAATCCTGTGAGATGAGTGTGCGAGAGAC 3498  
QY 648 TGGAGGTGCTCCAGTCTGCGAGGACAGTGTGCGAGAGTG 689  
DB 3499 GAGAGTGTGAGTGTGCTGCTGCGTCAACATCAAGAGAGCTG 3540

RESULT 95  
CQ583074 6124 bp DNA linear PAT 02-FEB-2004  
LOCUS CQ583074  
DEFINITION Sequence 10832 from Patent WO0171042.  
ACCESSION CQ583074  
VERSION CQ583074.1 GI:41644029  
KEYWORDS Drosophila sp.  
SOURCE Drosophila sp.  
ORGANISM Drosophila sp.

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	ORIGIN
1	Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.	Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more <i>Drosophila</i> genes and uses thereof	Patent: WO 0171042-A 10832 27-SEP-2001;	PE Corporation (NY) (US)	
2	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
3	Location/Qualifiers				
4	1. .6124				
5	/organism="Drosophila sp."				
6	/mol_type="unassigned DNA"				
7	/db_xref="taxon:7242"				
8	Query Match	3.7%; Score 56.4; DB 6; Length 6124;			
9	Best Local Similarity	59.3%; Pred. No. 0.6;			
10	Matches	96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;			
11	528 AATGAGCAGAGTGTGCGGTGTGCGGAGCGGAGAGCTCATCTGTGAGCGCTGC	587			
12	Db 3391 AACGAGACTGTGTGCGGCTGTCTGTGATAGGAGCGAGCTGATGTCTGCGACAATGT	3450			
13	Qy 588 CCTCGGCGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCACTGGAGCC	647			
14	Db 3451 CCCAAAGTTTCCATCAAGAACTGTCACTCCCTCGATCGATCGTTGCCGACGAGAC	3510			
15	Qy 648 TGGAGGTCTCCAGCTGCTGCGAGCAACATCCAGAGGCG	689			
16	Db 3511 GAGAGCTGGAGTGTCTATCTGTGCTGCTCAACTCAAGAGCTG	3552			
17	RESULT 96				
18	LOCUS	CO583073	21747 bp	DNA	Linear
19	DEFINITION	Sequence 10831 from Patent WO0171042.			PAT 02-FEB-2004
20	ACCESSION	CO583073			
21	VERSION	CO583073.1	GI:41644028		
22	KEYWORDS				
23	SOURCE				
24	ORGANISM				
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					

RESULT 97  
LOCUS AC017943 97538 bp DNA linear HTG 09-DEC-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
ACCESSION AC017943  
VERSION AC017943.1 GI:6553247  
KEYWORDS HTG, HTGS, PHASE2.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS Adams, M. and Venter, J.C.  
TITLE Direct Submission  
JOURNML Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT  
This sequence was identified as CDM:10212734 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
Source  
Location/Qualifiers  
1..97538  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
ORIGIN  
Query Match 3.7%; Score 56.4; DB 2; Length 97538;  
Best Local Similarity 59.3%; Pred. No. 0.38; Mismatches 66; Indels 0; Gaps 0;  
Matches 96; Conservative 0;  
QY 528 AATGAGCAGCATGTGCGGTGTGTCGGAGCGAGGAGCTCATCTGTGACGCGTC 587  
|||  
DB 9373 AACAGAGACTGTGCGCGCTGTCTGTGATGAGAGCGAGCTGATGCTGCGACAACTGT 9432  
588 CTTGGGCGCTTCCACCTGAGCTGCGCTGCTCCCTCCGCTCCGAGGATCCCACTGGAGCC 647  
DB 9433 CCAGAGTTTCCATCAGAACTGTCACATCCCTCGATCAGCTGTTGCCGAGAGAGC 9492  
648 TGGAGGTGCTCCAGCTGCTGCGAGCAACAGTCCAGAGGTG 689  
DB 9493 GAGAGCTGGCAGTCCCTACTGTGTGCTGCACATCAAGAGCTG 9534  
RESULT 98  
LOCUS AC007757 170939 bp DNA linear INV 15-FEB-2001  
DEFINITION Drosophila melanogaster, chromosome 3R, region 92E-92F, BAC clone  
AC007757  
ACCESSION AC007757  
VERSION AC007757  
KEYWORDS HTG, GI:12831350  
ORGANISM Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS Ceiniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,  
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
Rogers, Y., An, H., Baldwin, D., Bonzon, U., Beeson, K.Y., Busan, D.A.,  
Carlson, J.W., Center, A., Chapple, M., Davernport, L.B., Dietz, S.M.,  
Dodson, K., Dorey, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,  
Gonzalez, S., Frishe, E., Galle, R.F., Gary, N.S., George, R.A.,  
Renteria, M., Houck, J., Hoskins, R.A., Hostin, D., Howard, T.J.,  
Iibegam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moehref, A.,  
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
Paclet, J., Pargacs, V., Park, S., Patel, S., Pfeiffer, B.,

Phouanavong, S., Piltman, G.S., Purri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
Sequencing of Drosophila chromosome 3R, region 92B-92F  
Unpublished  
2 (bases 1 to 170939)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E., Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (07-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Feb 15, 2001 this sequence version replaced gi:6598748.

**COMMENT** Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

#### FEATURES

source  
1. 170939  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/strain="Y: cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
/map="92B-92F"  
/clone="BAC BACR13019 (D737)"  
/clone\_1ib="RPCI-98 (Roswell) Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBAC3.6")

#### ORIGIN

Query Match 3.7%; Score 56.4; DB 3; Length 170939;  
Best Local Similarity 59.3%; Pred. No. 0.35;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 528 AATGAGAGAGATGTCCTGTCGCGACGCGGGAGACTATCTGCTGTGACGGCTGC 587  
Db 131964 AACGAGAGACTGTGCGCCGCTGTCTGTGATGAGAGCGAGACTGATGTGTCGACAAAGTGT 132023  
QY 588 CCTCGGCGCTTCCACCTGCGCTGCTGCCCTCCGCTCCGGGAGATCCCGAGTGGAGCC 647  
Db 132024 CCCAAGTTTTCATCAGACATCTCACTCCCTGCGATCAGCTCGTTCGCGAGCAGAGCC 132083  
QY 648 TGGAGTGCTCCAGCTGCTCGACGAGCAAGTCCAGAGGTG 689  
Db 132084 GAGAGCTGGCAGTCTACTGTGCGCTCAACATCAAGAGAGCTG 132125

**RESULT 99**  
AC007771 197071 bp DNA linear INV 15-FEB-2001  
**LOCUS** Drosophila melanogaster, chromosome 3R, region 92B-92F, BAC clone  
**DEFINITION** BACR26M03, complete sequence.  
**ACCESSION** AC007771  
**VERSION** AC007771.7 GI:12831351  
**KEYWORDS** HTG.

**SOURCE** Drosophila melanogaster (fruit fly)  
**ORGANISM** Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

**REFERENCE** 1 (bases 1 to 197071)  
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Bazzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davaport, L.B., Dietz, S.M., Dodson, K., Dorselt, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreria, S., Frise, B., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibbagman, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunco, J., Pacled, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Pfeiffer, F., Phouanavong, S., Piltman, G.S., Purri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
Sequencing of Drosophila chromosome 3R, region 92B-92F  
Unpublished  
2 (bases 1 to 197071)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E., Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Feb 15, 2001 this sequence version replaced gi:7018763.

#### COMMENT

Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

#### FEATURES

source  
1. 197071  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/strain="Y: cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
/map="92E-92F"  
/clone="BAC BACR26M03 (D738)"  
/clone\_1ib="RPCI-98 (Roswell) Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBAC3.6")

#### ORIGIN

Query Match 3.7%; Score 56.4; DB 3; Length 197071;  
Best Local Similarity 59.3%; Pred. No. 0.34;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 528 AATGAGAGAGATGTCCTGTCGCGACGCGGGAGACTATCTGCTGTGACGGCTGC 587  
Db 19992 AACGAGAGACTGTGCGCCGCTGTCTGTGATGAGAGCGAGCTGATGCTGCGACAAAGTGT 20051  
QY 588 CCTCGGCGCTTCCACCTGCGCTGCTGCCCTCCGCTCCGGGAGATCCCGAGTGGAGCC 647  
Db 20052 CCCAAGTTTTCATCAGACATCTCACTCCCTGCGATCAGCTCGTTCGCGAGCAGAGCC 20111  
QY 648 TGGAGTGCTCCAGCTGCTCGACGAGCAAGTCCAGAGGTG 689  
Db 20112 GAGAGCTGGCAGTCTACTGTGCGCTCAACATCAAGAGAGCTG 20153

**RESULT 100**  
AE003731 230001 bp DNA linear INV 15-MAR-2004  
**LOCUS** Drosophila melanogaster chromosome 3R, section 69 of 118 of the  
**DEFINITION** complete sequence.  
**ACCESSION** AE003731 AE014297

VERSION	AE003731.3	GI:23175594
KEYWORDS	Drosophila melanogaster (fruit fly)	
SOURCE	Drosophila melanogaster	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 230001)	
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Mortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Chape,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabo,G.L., Ball,J.T.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,C., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolintineanu,D., Borkova,D., Botchan,M.R., Boulter,J., Brokstein,P., Brotter,P., Burks,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,J.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Diez,S.M., Dodson,K., Doup,L.E., Domes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriere,S., Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalish,F., Karpen,G.H., Ke,Z., Kentonson,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milbrandt,N.V., Mobarry,C., Mout,R.J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nuskern,D.R., Paclob,J.M., Palazzolo,M., Plattman,G.S., Pan,S., Pollard,J.D., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,P., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Maasarm,D.A., Weinstock,G.M., Weisenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.	
TITLE	The genome sequence of Drosophila melanogaster	
JOURNAL	Science 287 (5461), 2185-2195 (2000)	
MEDLINE	20196006	
PUBMED	10731132	
REFERENCE	2 (bases 1 to 230001)	
AUTHORS	Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W., Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frisze,E., Hodson,A., George,R.A., Hoskins,R.A., Laverly,T., Muzny,D.M., Nelson,C.R., Paclob,J.M., Park,S., Pfeiffer,B.D., Richards,S., Sodergren,E.J., Svirskas,R., Taboy,P.B., Wan,K., Stapleton,M., Sutton,G.G., Venter,C., Weinstock,G., Scherer,S.E., Myers,E.W., Gibbs,R.A. and Rubin,G.M.	
TITLE	Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence	
JOURNAL	Genome Biol. 3 (12), RESEARCH0079 (2002)	
MEDLINE	22426065	
PUBMED	12537568	
REFERENCE	3 (bases 1 to 230001)	
AUTHORS	Mista,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E., Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L., Berman,B.P., Betancourt,B.R., Celniker,S.E., de Grey,A.D., Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J., Shu,S.Q., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M., Rubin,G.M. and Lewis,S.E.	
TITLE	Annotation of the Drosophila melanogaster euchromatic genome: a systematic review	
JOURNAL	Genome Biol. 3 (12), RESEARCH0083 (2002)	
MEDLINE	22426069	
PUBMED	12537572	
REFERENCE	4 (bases 1 to 230001)	
AUTHORS	Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J., Svirskas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E.	
TITLE	The transposable elements of the Drosophila melanogaster euchromatin: a genomic perspective	
JOURNAL	Genome Biol. 3 (12), RESEARCH0084 (2002)	
MEDLINE	22426070	
PUBMED	12537573	
REFERENCE	5 (bases 1 to 230001)	
AUTHORS	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA	
REFERENCE	6 (bases 1 to 230001)	
AUTHORS	FlyBase	
CONSRMT	Direct Submission	
REFERENCE	Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA	
CONSRMT	7 (bases 1 to 230001)	
TITLE	FlyBase	
CONSRMT	Direct Submission	
REFERENCE	Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA	
CONSRMT	On Sep 18, 2002 this sequence version replaced gi:10726652.	
FEATURES	Location/Qualifiers	
source	1..230001	
gene	/organism="Drosophila melanogaster"	
gene	/mol_type="genomic DNA"	
gene	/db_xref="taxon:7227"	
gene	/chromosome="3R"	
gene	11240..11520	
gene	/gene="MtnD"	
gene	/locus_tag="CG33192"	
gene	/map="92F1..92F1"	
gene	/db_xref="FLYBASE:FBgn0053192"	
gene	join(11240..11276,11343..11520)	
gene	/gene="MtnD"	
gene	/locus_tag="CG33192"	
gene	/product="CG33192-RA"	
gene	/db_xref="FLYBASE:FBgn0053192"	
gene	join(11252..11276,11343..11433)	
gene	/gene="MtnD"	
gene	/locus_tag="CG33192"	
gene	/codon_start=1	
gene	/product="CG33192-PA"	
gene	/protein_id="AAO41579.1"	
gene	/db_xref="GI:20381377"	
gene	/db_xref="FLYBASE:FBgn0053192"	
gene	/translation="MGCKACGNCGCSATKCGDNCAGSQCCGSC"	
gene	complement(11918..28547)	
gene	/gene="Stat92E"	
gene	/locus_tag="CG4257"	
gene	/map="92F1..92F1"	
gene	/db_xref="FLYBASE:FBgn0016917"	
gene	complement(join(11918..12760,13429..13698,15456..15570,15635..15835,15894..16151,16210..16689,16758..17788,17915..18327,28316..28547))	
gene	/gene="Stat92E"	
gene	/locus_tag="CG4257"	
gene	/product="CG4257-RE"	
gene	/note="This object was named CG4257-RD in FlyBase D.melanogaster release 3.1"	
gene	/db_xref="FLYBASE:FBgn0016917"	
gene	complement(join(11918..12760,13429..13698,15456..15591,15635..15835,15894..16151,16210..16689,16758..17788,17915..18327,28316..28547))	
gene	/gene="Stat92E"	
gene	/locus_tag="CG4257"	
gene	/locus_tag="CG4257"	





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 739 Seconds

(without alignment)  
10974.771 Million cell updates/sec

Title: US-09-508-658a-3

Perfect score: 1545  
Sequence: 1 agaggaagctgaggtcttcctc.....aaaataataaactgctg 1545

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 200 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: geneseqn1808:\*  
2: geneseqn1808:\*  
3: geneseqn20008:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541.8	99.8	1545	2	AAX26937
2	1367.8	88.5	1463	2	AAX26938
3	1023.6	66.3	2036	2	AAX26936
4	1023.6	66.3	2245	2	AAX26935
5	385.6	25.0	1656	2	AAX58606
6	118.4	7.7	9876	6	ABLJ2092
7	101.8	6.6	9876	6	ABLJ2093
8	4.2	9646	12	ADH12179	
9	3.9	60	6	ABN38615	
10	60	5739	6	ABK70271	
11	60	6327	6	AAI32301	
12	60	6417	6	ABL64410	
13	60	6417	6	ABX04169	
14	60	6417	6	ABK84696	
15	60	6417	10	AAAD54631	
16	60	6417	10	AAAD54632	
17	60	6417	12	ADP10531	
18	60	6475	4	AAI58444	
19	60	6475	5	ADQ98655	
20	60	6475	9	ADBA8415	
21	60	6553	4	AAI60230	

22	60	3.9	6648	12	ADQ22446	Adq22446 Human scf
23	59.8	3.9	1053	2	AAQ22728	AAQ22728 M1-2 anti
24	59.8	3.9	1053	2	ADP69743	ADP69743 cDNA enco
25	59.8	3.9	6331	5	ABV23250	ABV23250 Human pro
26	59.8	3.9	6331	8	ACD13374	ACD13374 Human DNA
27	59.8	3.9	6331	10	AD126769	AD126769 Human CHD
28	59.8	3.9	6331	10	AD126768	AD126768 Human CHD
29	59.8	3.9	6331	10	AD546429	AD546429 Human chr
30	59.8	3.9	6771	10	AD126742	AD126742 Human GS-
31	59.8	3.9	7805	8	ABX63370	ABX63370 Human cDN
32	58.8	3.7	7805	8	AD122590	AD122590 Human liv
33	58.8	3.7	2000	8	ADA71938	ADA71938 Rice gene
34	57.4	3.7	114955	2	AAV53491	AAV53491 Human ade
35	56.4	3.7	6124	4	ABL09061	ABL09061 Drosophi
36	56.4	3.7	2147	4	ABL09060	ABL09060 Human gen
37	55.4	3.6	153	12	ACH90409	ACH90409 Human gen
38	55.4	3.6	559	12	ACH76709	ACH76709 Human gen
39	55.4	3.6	2495	5	AA582330	AA582330 DNA enco
40	55.4	3.6	7132	5	AA574874	AA574874 DNA enco
41	55.4	3.6	37314	4	AAK71358	AAK71358 Human ltm
42	55.2	3.6	172570	6	ABQ88207	ABQ88207 Human ost
43	55.2	3.6	523	3	AAA45365	AAA45365 Mouse sec
44	54.2	3.5	494	9	ACH48320	ACH48320 Human lun
45	54.2	3.5	3039	9	ADB80995	ADB80995 RING-SH C
46	54.2	3.5	3917	10	ADJ56478	ADJ56478 Human cDN
47	54.2	3.5	3997	4	AAK94864	AAK94864 Human ful
48	54.2	3.5	3997	12	ADL32011	ADL32011 Full leng
49	54	3.5	114955	2	AAV53491	AAV53491 Human ade
50	52.8	3.4	292	4	AAI25090	AAI25090 Probe #15
51	52.8	3.4	292	4	ABA70776	ABA70776 Human foe
52	52.8	3.4	292	4	AAI50949	AAI50949 Probe #19
53	52.8	3.4	292	4	ABA37266	ABA37266 Probe #15
54	52.8	3.4	292	4	AAK44980	AAK44980 Human bon
55	52.8	3.4	292	4	AAK19029	AAK19029 Human bra
56	52.8	3.4	292	4	ABE44646	ABE44646 Human liv
57	52.8	3.4	292	6	ABE19225	ABE19225 Human gen
58	52.8	3.4	476	4	AAI15873	AAI15873 Probe #58
59	52.8	3.4	476	4	ABA58175	ABA58175 Human foe
60	52.8	3.4	476	4	AAI37782	AAI37782 Probe #64
61	52.8	3.4	476	4	ABA27360	ABA27360 Probe #58
62	52.8	3.4	476	4	AAK31915	AAK31915 Human bon
63	52.8	3.4	476	4	AAK06255	AAK06255 Human bra
64	52.8	3.4	476	4	ABE31605	ABE31605 Human liv
65	52.8	3.4	476	6	ABE06677	ABE06677 Human gen
66	52.8	3.4	6025	4	ABL29649	ABL29649 Drosophi
67	52.8	3.4	7303	4	ABL29648	ABL29648 Drosophi
68	52.6	3.4	4653	6	ABQ99469	ABQ99469 Human cod
69	52.2	3.4	539	5	ABV44220	ABV44220 Human pro
70	52.2	3.4	539	5	ABV29094	ABV29094 Human pro
71	52	3.4	1670	10	ADC30621	ADC30621 Human nov
72	52	3.4	1651	10	AAK99602	AAK99602 MDTT rela
73	52	3.4	2651	11	ADM02221	ADM02221 Human cDN
74	52	3.4	3498	10	ADB62798	ADB62798 Human cDN
75	52	3.4	3584	10	ADB62384	ADB62384 Human cDN
76	52	3.4	3710	12	ADQ24519	ADQ24519 Human bof
77	50.6	3.3	2000	8	ADK71238	ADK71238 Rice gene
78	48.8	3.2	2226	12	ADU37277	ADU37277 Human nuc
79	48.6	3.2	3079	4	AAH16238	AAH16238 Human cDN
80	48.6	3.1	2171	4	AAK52998	AAK52998 Human pol
81	48.6	3.1	2179	4	AAK52014	AAK52014 Human pol
82	48	3.1	3005	12	ADQ35478	ADQ35478 Novel mou
83	47.8	3.1	924	5	AA575442	AA575442 DNA enco
84	47.8	3.1	1006	4	AAH33830	AAH33830 Human col
85	47.8	3.1	2724	4	AAH16455	AAH16455 Human cDN
86	47.8	3.1	3424	12	ADP13566	ADP13566 Renal cel
87	47.8	3.1	3510	9	ADB80996	ADB80996 RING-SH C
88	47.8	3.1	6730	11	ADL22564	ADL22564 Human dis
89	47.6	3.1	3323	4	ABU05755	ABU05755 Drosophi
90	47.2	3.1	2805	2	AAV24559	AAV24559 Leukocyte
91	47.2	3.1	2911	5	AA575444	AA575444 DNA enco
92	47.2	3.1	3352	5	AA575443	AA575443 DNA enco
93	47	3.0	1210	5	AA575445	AA575445 DNA enco
94	47	3.0	6610	4	AA529014	AA529014 Genomic s

95	47	3.0	6610	4	AA29013	AA29013 Genomic s
96	47	3.0	6610	4	AA26967	AA26967 Human gen
97	47	3.0	6610	4	AA26968	AA26968 Human gen
98	47	3.0	6610	4	AA26969	AA26969 Human gen
99	47	3.0	6610	4	AA26970	AA26970 Human gen
100	47	3.0	6610	6	AA26978	AA26978 Human gen
101	47	3.0	6610	6	AA26978	AA26978 Human gen
102	47	3.0	6610	6	AA26978	AA26978 Human gen
103	47	3.0	6610	6	AA26978	AA26978 Human gen
104	47	3.0	6610	6	AA26978	AA26978 Human gen
105	46.6	3.0	6610	6	AA26978	AA26978 Human gen
106	46.6	3.0	6610	6	AA26978	AA26978 Human gen
107	46	3.0	6610	6	AA26978	AA26978 Human gen
108	45.8	3.0	6610	6	AA26978	AA26978 Human gen
109	45.4	2.9	6610	6	AA26978	AA26978 Human gen
110	45	2.9	6610	6	AA26978	AA26978 Human gen
111	45	2.9	6610	6	AA26978	AA26978 Human gen
112	45	2.9	6610	6	AA26978	AA26978 Human gen
113	45	2.9	6610	6	AA26978	AA26978 Human gen
114	44.6	2.9	6610	6	AA26978	AA26978 Human gen
115	44.6	2.9	6610	6	AA26978	AA26978 Human gen
116	44.6	2.9	6610	6	AA26978	AA26978 Human gen
117	44.6	2.9	6610	6	AA26978	AA26978 Human gen
118	44.6	2.9	6610	6	AA26978	AA26978 Human gen
119	44.6	2.9	6610	6	AA26978	AA26978 Human gen
120	44.2	2.9	6610	6	AA26978	AA26978 Human gen
121	44.2	2.9	6610	6	AA26978	AA26978 Human gen
122	44	2.8	6610	6	AA26978	AA26978 Human gen
123	44	2.8	6610	6	AA26978	AA26978 Human gen
124	44	2.8	6610	6	AA26978	AA26978 Human gen
125	43.8	2.8	6610	6	AA26978	AA26978 Human gen
126	43.8	2.8	6610	6	AA26978	AA26978 Human gen
127	43.8	2.8	6610	6	AA26978	AA26978 Human gen
128	43.8	2.8	6610	6	AA26978	AA26978 Human gen
129	43.8	2.8	6610	6	AA26978	AA26978 Human gen
130	43.8	2.8	6610	6	AA26978	AA26978 Human gen
131	43.6	2.8	6610	6	AA26978	AA26978 Human gen
132	43.6	2.8	6610	6	AA26978	AA26978 Human gen
133	43.6	2.8	6610	6	AA26978	AA26978 Human gen
134	43.6	2.8	6610	6	AA26978	AA26978 Human gen
135	43.4	2.8	6610	6	AA26978	AA26978 Human gen
136	43.4	2.8	6610	6	AA26978	AA26978 Human gen
137	43.4	2.8	6610	6	AA26978	AA26978 Human gen
138	43.4	2.8	6610	6	AA26978	AA26978 Human gen
139	43.4	2.8	6610	6	AA26978	AA26978 Human gen
140	43.2	2.8	6610	6	AA26978	AA26978 Human gen
141	43	2.8	6610	6	AA26978	AA26978 Human gen
142	43	2.8	6610	6	AA26978	AA26978 Human gen
143	43	2.8	6610	6	AA26978	AA26978 Human gen
144	43	2.8	6610	6	AA26978	AA26978 Human gen
145	43	2.8	6610	6	AA26978	AA26978 Human gen
146	43	2.8	6610	6	AA26978	AA26978 Human gen
147	43	2.8	6610	6	AA26978	AA26978 Human gen
148	43	2.8	6610	6	AA26978	AA26978 Human gen
149	43	2.8	6610	6	AA26978	AA26978 Human gen
150	43	2.8	6610	6	AA26978	AA26978 Human gen
151	42.8	2.8	6610	6	AA26978	AA26978 Human gen
152	42.8	2.8	6610	6	AA26978	AA26978 Human gen
153	42.8	2.8	6610	6	AA26978	AA26978 Human gen
154	42.8	2.8	6610	6	AA26978	AA26978 Human gen
155	42.8	2.8	6610	6	AA26978	AA26978 Human gen
156	42.8	2.8	6610	6	AA26978	AA26978 Human gen
157	42.8	2.8	6610	6	AA26978	AA26978 Human gen
158	42.8	2.8	6610	6	AA26978	AA26978 Human gen
159	42.8	2.8	6610	6	AA26978	AA26978 Human gen
160	42.8	2.8	6610	6	AA26978	AA26978 Human gen
161	42.8	2.8	6610	6	AA26978	AA26978 Human gen
162	42.8	2.8	6610	6	AA26978	AA26978 Human gen
163	42.8	2.8	6610	6	AA26978	AA26978 Human gen
164	42.8	2.8	6610	6	AA26978	AA26978 Human gen
165	42.8	2.8	6610	6	AA26978	AA26978 Human gen
166	42.6	2.8	6610	6	AA26978	AA26978 Human gen
167	42.6	2.8	6610	6	AA26978	AA26978 Human gen
168	42.6	2.8	6610	6	AA26978	AA26978 Human gen
169	42.4	2.7	6610	6	AA26978	AA26978 Human gen
170	42.4	2.7	6610	6	AA26978	AA26978 Human gen
171	42.4	2.7	6610	6	AA26978	AA26978 Human gen
172	42.2	2.7	6610	6	AA26978	AA26978 Human gen
173	42.2	2.7	6610	6	AA26978	AA26978 Human gen
174	42.2	2.7	6610	6	AA26978	AA26978 Human gen
175	42.2	2.7	6610	6	AA26978	AA26978 Human gen
176	42.2	2.7	6610	6	AA26978	AA26978 Human gen
177	42.2	2.7	6610	6	AA26978	AA26978 Human gen
178	42.2	2.7	6610	6	AA26978	AA26978 Human gen
179	42.2	2.7	6610	6	AA26978	AA26978 Human gen
180	42.2	2.7	6610	6	AA26978	AA26978 Human gen
181	42.2	2.7	6610	6	AA26978	AA26978 Human gen
182	42.2	2.7	6610	6	AA26978	AA26978 Human gen
183	42.2	2.7	6610	6	AA26978	AA26978 Human gen
184	42.2	2.7	6610	6	AA26978	AA26978 Human gen
185	42	2.7	6610	6	AA26978	AA26978 Human gen
186	41.8	2.7	6610	6	AA26978	AA26978 Human gen
187	41.6	2.7	6610	6	AA26978	AA26978 Human gen
188	41.6	2.7	6610	6	AA26978	AA26978 Human gen
189	41.4	2.7	6610	6	AA26978	AA26978 Human gen
190	41.4	2.7	6610	6	AA26978	AA26978 Human gen
191	41.4	2.7	6610	6	AA26978	AA26978 Human gen
192	41.4	2.7	6610	6	AA26978	AA26978 Human gen
193	41.4	2.7	6610	6	AA26978	AA26978 Human gen
194	41.2	2.7	6610	6	AA26978	AA26978 Human gen
195	41.2	2.7	6610	6	AA26978	AA26978 Human gen
196	41.2	2.7	6610	6	AA26978	AA26978 Human gen
197	41.2	2.7	6610	6	AA26978	AA26978 Human gen
198	41.2	2.7	6610	6	AA26978	AA26978 Human gen
199	41	2.7	6610	6	AA26978	AA26978 Human gen
200	41	2.7	6610	6	AA26978	AA26978 Human gen
ALIGNMENTS						
RESULT 1						
AAX26937						
ID	AAX26937	etandard;	DNA;	1545	BP.	
AAX26937;						
24-JUN-1999 (first entry)						
cDNA encoding a human autoimmune regulator-2 (AIR-2) protein.						
Autoimmune regulator-2; AIR-2; immune maturation; immune response;						
disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;						
APECED; autoimmune polyglandular syndrome type I; APS I; ss.						
Homo sapiens.						
Key Location/Qualifiers						
FT	CDS	237..1283				
FT	/*tag= a					
FT	/product= "AIR-2"					
W09915559-A1.						
01-APR-1999.						
23-SEP-1997; 97FI-00003762.						
(FIM-) FINNISH IMMUNOTECHNOLOGY LTD.						
Krohn K, Heino M, Peterson P, Scott H, Antonarakis S, Lalioti M;						
Shimizu N, Kudoh J;						
PI Shinzou N, Kidoh J;						
WPI; 1999-244390/20.						



PN WO9915559-A1.  
XX 01-APR-1999.  
XX 23-SEP-1998; 98WO-FI000749.  
XX 23-SEP-1997; 97FI-00003762.  
XX (FIM-) FINNISH IMMUNOTECHNOLOGY LTD.  
XX Krohn K, Heino M, Peterson P, Scott H, Antonarakis S, Lallioi M,  
XX Shimizu N, Kudoh J;  
XX MPI; 1999-244390/20.  
XX P-PSDB; AAY01714.  
XX  
XX Autoimmune regulator 1 (AIR1) DNA sequence.  
XX  
XX Claim 4; Page 31-32; 59pd; English.  
XX  
XX The present sequence encodes an autoimmune regulator-3 (AIR-3) protein.  
XX The AIR polypeptides and polynucleotides can be used in methods for the  
XX diagnosis and treatment of diseases related to immune maturation and  
XX regulation of immune response towards self and nonself. They can be used  
XX particularly in the diagnosis and treatment of autoimmune  
XX polyendocrinopathy candidiasis ectodermal dysplasia (APECED) (also known  
XX as autoimmune polyglandular syndrome type I (APS I))

Sequence 1463 BP; 245 A; 456 C; 486 G; 276 T; 0 U; 0 Other;

Query Match 88.5%; Score 1367.8; DB 2; Length 1463;  
Best Local Similarity 94.6%; Pred. No. 4.1e-282;  
Matches 1461; Conservative 0; Mismatches 2; Indels 82; Gaps 1;

QY 1 AAGAGAAAGGAGGTCTTCTCAGGCTTTAAGAGCATGGGCTTGTTCAGGCTGTACCG 60  
DB 1 AAGAGAAAGGAGGTCTTCTCAGGCTTTAAGAGCATGGGCTTGTTCAGGCTGTACCG 60  
QY 61 CTGCTCTCAGCTGGGCGCCGTGGGTGGGCGGGGCGCCCTGCTATAGCCAGAGGTCAAG 120  
DB 61 CTGCTCTCAGCTGGGCGCCGTGGGTGGGCGGGGCGCCCTGCTATAGCCAGAGGTCAAG 120  
QY 121 ATCCACTGGGAATGCCATGCTATCTTTCGTCGCCAGATGTTTCTTAAATGGGATAGAA 180  
DB 121 ATCCACTGGGAATGCCATGCTATCTTTCGTCGCCAGATGTTTCTTAAATGGGATAGAA 180  
QY 181 GCAAGTCGGGAGAGACTCTCTGGGCTTGAGCCCACTGCTGTAGAGAAAGGTTATGT 240  
DB 181 GCAAGTCGGGAGAGACTCTCTGGGCTTGAGCCCACTGCTGTAGAGAAAGGTTATGT 240  
QY 241 GGTGGGTGTACAGTTCGGGGGCGCCCTGGAACGAGAGAGCTGTCAAGAAACCGGGTTTCT 300  
DB 241 GGTGGGTGTACAGTTCGGGGGCGCCCTGGAACGAGAGAGCTGTCAAGAAACCGGGTTTCT 300  
QY 301 TCCCAATAGGATGAGCCCGGGGGGTGTCTGTTGAGACCAATGATGGGAAACAGGTG 360  
DB 301 TCCCAATAGGATGAGCCCGGGGGGTGTCTGTTGAGACCAATGATGGGAAACAGGTG 360  
QY 361 GTCAAGGGCAGAAATTTAGAGCCCTTGAGAGATAGGAGCAGAGATCGGGAGTTCAAG 420  
DB 361 GTCAAGGGCAGAAATTTAGAGCCCTTGAGAGATAGGAGCAGAGATCGGGAGTTCAAG 420  
QY 421 GTACCCAGAGATGCTCTGGGGAGCTGTTTGGAAAGAGAGTGGCTCTCAGAGAGGTGC 480  
DB 421 GTACCCAGAGATGCTCTGGGGAGCTGTTTGGAAAGAGAGTGGCTCTCAGAGAGGTGC 480  
QY 481 TGCACCCAGCCAGCTGTGATGAGGCGTCTTTCCTGTGCCAGAAATGAGACGAGT 540  
DB 481 TGCACCCAGCCAGCTGTGATGAGGCGTCTTTCCTGTGCCAGAAATGAGACGAGT 540  
QY 541 GTGCCGTGTGTGGGAGCGGCGGGAGACTCATCTGTGTGACGAGCTGCCCTCGGGCTTCC 600  
DB 541 GTGCCGTGTGTGGGAGCGGCGGGAGACTCATCTGTGTGACGAGCTGCCCTCGGGCTTCC 600

QY 601 ACTGGGCGGCGTGTCCCTCCGCTCCGGAGATCCCAATGGGACCTGGAGGTCTCCA 660  
DB 601 ACTGGGCGGCGTGTCCCTCCGCTCCGGAGATCCCAATGGGACCTGGAGGTCTCCA 660  
QY 661 GCTGCTTGAGGCAACAGTCCAGAGAGTGCAGCCCGGGCAGAGAGCCCGGCCAGG 720  
DB 661 GCTGCTTGAGGCAACAGTCCAGAGAGTGCAGCCCGGGCAGAGAGCCCGGCCAGG 720  
QY 721 AGCCACCCGTGGAGAACCCGCTTCCCGGGGCTTAGGTTCGGCGGAGAGAGGTAAAG 780  
DB 721 AGCCACCCGTGGAGAACCCGCTTCCCGGGGCTTAGGTTCGGCGGAGAGAGGTAAAG 780  
QY 781 GTCCACTGGGGAACCCCTAGCCGAGATGACACGACTTGTGTCAAGACACTGCGG 840  
DB 781 GTCCACTGGGGAACCCCTAGCCGAGATGACACGACTTGTGTCAAGACACTGCGG 840  
QY 841 CTCCGCTTCTGACAGCCCGCTGCGAGATCTGAGACTCTCGGCTTGACACCCCTACTGT 900  
DB 841 CTCCGCTTCTGACAGCCCGCTGCGAGATCTGAGACTCTCGGCTTGACACCCCTACTGT 900  
QY 901 GTGTGGTCTTGAGAGGTTCAGAGAACTTGGCTTGTGTGCGGCTTGGGGGTGTGGAG 960  
DB 901 GTGTGGTCTTGAGAGGTTCAGAGAACTTGGCTTGTGTGCGGCTTGGGGGTGTGGAG 960  
QY 961 ATGGTACGAGACGTCGCGGTGATCACTGCGCGCTGCTTCCACTGAGCGTGCACCT 1020  
DB 961 ATGGTACGAGACGTCGCGGTGATCACTGCGCGCTGCTTCCACTGAGCGTGCACCT 1020  
QY 1021 TCCACCCGCGACCTCCGCGCCCGGAGCGGAGCTGGCTGACAGATCTGCTCAGAGAGC 1080  
DB 1021 TCCACCCGCGACCTCCGCGCCCGGAGCGGAGCTGGCTGACAGATCTGCTCAGAGAGC 1080  
QY 1081 TGAACCCAGCCCTGTGAGAGGGGTGTGCGCCCGAGCCCGCCCGGCTGGGCGCTGGG 1140  
DB 1081 TGAACCCAGCCCTGTGAGAGGGGTGTGCGCCCGAGCCCGCCCGGCTGGGCGCTGGG 1140  
QY 1141 CTGCCAAGATGACACTGCACTGACAGAGCCGCTCTGACAGAGATGACCTGAGTCCC 1200  
DB 1141 CTGCCAAGATGACACTGCACTGACAGAGCCGCTCTGACAGAGATGACCTGAGTCCC 1200  
QY 1201 TTCTGAGCAGACACCTTGTGATGAGCTCTGAGAGTGGGCTGACAGAGATGAGCCGCT 1260  
DB 1201 TTCTGAGCAGACACCTTGTGATGAGCTCTGAGAGTGGGCTGACAGAGATGAGCCGCT 1260  
QY 1261 CGGCGGCGCCCTTCCCTCCTGACCCAGATGGCGGAGCATGAGAGAG 1320  
DB 1261 CGGCGGCGCCCTTCCCTCCTGACCCAGATGGCGGAGCATGAGAGAG 1320  
QY 1321 TGTGAGAGAGACACTCTCTCTCTCAGTCTGGAAGCCGCGGCTGGGATCAAGAGG 1380  
DB 1321 TGTGAGAGAGACACTCTCTCTCTCAGTCTGGAAGCCGCGGCTGGGATCAAGAGG 1380  
QY 1381 GACAGCGCACCTTGTGATGAGTCTGCTGTAACAGCTGTGTGTTTCTGGGACACCA 1440  
DB 1381 GACAGCGCACCTTGTGATGAGTCTGCTGTAACAGCTGTGTGTTTCTGGGACACCA 1440  
QY 1441 GGCATGATGCTGGAATTAACCCCTGCCCACTTCTCTCTGGAAGTCCCGGGA 1500  
DB 1441 GGCATGATGCTGGAATTAACCCCTGCCCACTTCTCTCTGGAAGTCCCGGGA 1500  
QY 1501 GCCTCTCTGCTGTGATCACTTAATAATTAATAATTAAGCTG 1545  
DB 1501 GCCTCTCTGCTGTGATCACTTAATAATTAATAATTAAGCTG 1545  
QY 1419 GCCTCTCTGCTGTGATCACTTAATAATTAATAATTAAGCTG 1463  
DB 1419 GCCTCTCTGCTGTGATCACTTAATAATTAATAATTAAGCTG 1463

RESULT 3  
AAAX26936  
ID AAAX26936 standard; DNA; 2036 BP.  
XX  
AC AAAX26936;  
XX  
DT 24-JUN-1999 (first entry)

XX CDNA encoding a human autoimmune regulator-1 (AIR-1) protein.  
XX  
XX Autoimmune regulator-1; AIR-1; Immune maturation; immune response;  
XX disease; autoimmune polyendocrinopathy candidiasis ectodermal dys trophy;  
XX APECED; autoimmune polyglandular syndrome type I; APS I; ss.  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 137..1774  
XX FT /\*tag= a  
XX FT /product= "AIR-1"  
XX  
XX WO915559-A1.  
XX  
XX 01-APR-1999.  
XX  
XX 23-SEP-1998; 98WO-FI000749.  
XX  
XX 23-SEP-1997; 97FI-00003762.  
XX  
XX (FIM-) FINNISH IMMUNOTECHNOLOGY LTD.  
XX  
XX Krohn K, Heino M, Peterson P, Scott H, Antonarakis S, Lalioti M,  
XX Shmizu N, Kudo J;  
XX  
XX WPI; 1999-244390/20.  
XX  
XX P-PSDB; AAY01712.  
XX  
XX Autoimmune regulator 1 (AIR1) DNA sequence.  
XX  
XX Claim 4; Page 21-24; 59pp; English.  
XX  
XX The present sequence encodes an autoimmune regulator-1 (AIR-1) protein.  
XX The AIR polypeptides and polynucleotides can be used in methods for the  
XX diagnosis and treatment of diseases related to immune maturation and  
XX regulation of immune response towards self and nonself. They can be used  
XX particularly in the diagnosis and treatment of autoimmune  
XX polyendocrinopathy candidiasis ectodermal dys trophy (APECED) (also known  
XX as autoimmune polyglandular syndrome type I (APS I))  
XX  
XX Sequence 2036 BP; 359 A; 718 C; 649 G; 310 T; 0 U; 0 Other;  
SQ  
Query Match 66.3%; Score 1023.6; DB 2; Length 2036;  
Best Local Similarity 99.6%; Pred. No. 1.1e-208;  
Matches 1026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 516 CTGTGCGCAAGAAATGAGAGCGAGTGTGCGGTGTCGGGACGCGGGAGCTCATCTGC 575  
DB 1007 CTCACACGAAGAATGAGAGCGAGTGTGCGGTGTCGGGACGCGGGAGCTCATCTGC 1066  
QY 576 TGTGACGGCTGCGCTCGGGCTTCCACCTGAGCTGCTGCTCCGCTCCGGAGATC 635  
DB 1067 TGTGACGGCTGCGCTCGGGCTTCCACCTGAGCTGCTGCTCCGCTCCGGAGATC 1126  
QY 636 CCCAGTGGAGCTGAGGTGCTTCAGCTGCTGAGGCAACAGTCCAGAGGTGACGCC 695  
DB 1127 CCCAGTGGAGCTGAGGTGCTTCAGCTGCTGAGGCAACAGTCCAGAGGTGACGCC 1186  
QY 696 CGGGAGAGAGAGCCCGGGCCCGAGGACACCCGTGAGACCCCGCTCCCGGGGGCTT 755  
DB 1187 CGGGAGAGAGAGCCCGGGCCCGAGGACACCCGTGAGACCCCGCTCCCGGGGGCTT 1246  
QY 756 AGGTGCGCGGAGAGAGGTGAGAGGTTCACCTGAGGAAACCTTACCGGCATGACAG 815  
DB 1247 AGGTGCGCGGAGAGAGGTGAGAGGTTCACCTGAGGAAACCTTACCGGCATGACAG 1306  
QY 816 ACTCTTGTCTACAGACCTGCGGCTCGGCTTCTGACAGCCCGCTGCCAGGTCTGAC 875  
DB 1307 ACTCTTGTCTACAGACCTGCGGCTCGGCTTCTGACAGCCCGCTGCCAGGTCTGAC 1366  
QY 876 TCTGTGCGCCTGACACCCCTACTGTGTGTGGGTCTGAGAGGTGACAGAACTGTGCTCT 935

DB 1367 TCTGTGCGCCTGACACCCCTACTGTGTGTGGGTCTGAGAGGTGACAGAACTGTGCTCT 1426  
QY 936 GGTGCGCCTTGGCGGGGTGTGCGAGATGTAACGACGTCGCTGCTGTACTGACGCGC 995  
DB 1427 GGTGCGCCTTGGCGGGGTGTGCGAGATGTAACGACGTCGCTGCTGTACTGACGCGC 1486  
QY 996 GCTGCTTCTCACTGGGCTGCACTTCCAGCGGCACTCCCGCCGGGACGCGGCTG 1055  
DB 1487 GCTGCTTCTCACTGGGCTGCACTTCCAGCGGCACTCCCGCCGGGACGCGGCTG 1546  
QY 1056 CGCTGAGATCTGTGCTCAGAGACGTGACCCAGCCCGCTGAGAGGGGGTGTGCGCCCC 1115  
DB 1547 CGCTGAGATCTGTGCTCAGAGACGTGACCCAGCCCGCTGAGAGGGGGTGTGCGCCCC 1606  
QY 1116 AGCCCGCCGCGCTGAGCCCTGAGGCTGCGCAAGATGACATGCGACGACGCGCT 1175  
DB 1607 AGCCCGCCGCGCTGAGCCCTGAGGCTGCGCAAGATGACATGCGACGACGCGCT 1666  
QY 1176 CTGACAGGAGATGACCTGAGATCTCTTCTGAGCGACACACCTTGATGGCATCTGAC 1235  
DB 1667 CTGACAGGAGATGACCTGAGATCTCTTCTGAGCGACACACCTTGATGGCATCTGAC 1726  
QY 1236 TGGGCGATCCAGAGATGCGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1295  
DB 1727 TGGGCGATCCAGAGATGCGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1786  
QY 1296 GGGACATGAGCTGTGATGAGAGTGTGCTGAGAGACACCTTCTCTCACTGAGTGA 1355  
DB 1787 GGGACATGAGCTGTGATGAGAGTGTGCTGAGAGACACCTTCTCTCACTGAGTGA 1846  
QY 1356 GCGGCGCGGCTGGGATCAAGAGGAGACAGCGCCACCTTGTCACTGCTCGGCTTAA 1415  
DB 1847 GCGGCGCGGCTGGGATCAAGAGGAGACAGCGCCACCTTGTCACTGCTCGGCTTAA 1906  
QY 1416 CAGCTGTGTTCTGCGGAGACACGACCATGATGCTGAGAAATTAACCTGCGCCAC 1475  
DB 1907 CAGCTGTGTTCTGCGGAGACACGACCATGATGCTGAGAAATTAACCTGCGCCAC 1966  
QY 1476 TTCTACTCTGGAAGTCCCGGAGACCTCTGCTGCTGAGTGAACCTTAATAATA 1535  
DB 1967 TTCTACTCTGGAAGTCCCGGAGACCTCTGCTGCTGAGTGAACCTTAATAATA 2026  
QY 1536 AAATTAGCTG 1545  
DB 2027 AAATTAGCTG 2036  
RESULT 4  
AAK58605  
ID AAK58605 standard; cDNA; 2245 BP.  
XX  
XX AAK58605;  
XX  
XX 16-AUG-1999 (first entry)  
XX  
XX Human autoimmune polyglandular disease type 1 (APGD1) CDNA.  
XX  
XX Autoimmune polyendocrinopathy candidiasis ectodermal dys trophy; APECED;  
XX autoimmune polyglandular disease type 1; APGD1; AIRB; human;  
XX transcrip tion factor; autoimmune disease; diagnosis; gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 121..1758  
XX FT /\*tag= a  
XX FT /product= "human\_APGD1"  
XX FT replace(358,T)  
XX FT /\*tag= k  
XX FT mutation  
XX FT replace(374,G)  
XX FT /\*tag= 1  
XX FT variation  
XX replace(708,T)

Accession	Gene	Protein	Protein Description
FT		/*tag= o	"silent polymorphism"
FT		/note= "silent polymorphism"	
FT	variation	replace(801,T)	
FT		/*tag= p	"silent polymorphism"
FT	mutation	replace(889,T)	
FT		/*tag= j	"alters Arg codon to STOP, and truncated 256 aa protein"
FT		/note= "alters Arg codon to STOP, and truncated 256 aa protein"	
FT	mutation	1051	
FT		/*tag= g	"1-nucleotide deletion"
FT	mutation	replace(1052,A)	
FT		/*tag= m	
FT	mutation	1085..1097	
FT		/*tag= f	"13-nucleotide deletion, results in frameshift and truncated 372 aa protein"
FT	mutation	replace(1085,CTTG)	
FT		/*tag= b	"4-nucleotide insertion, results in frameshift and truncated 371 aa protein"
FT	mutation	/note= "4-nucleotide insertion, results in frameshift and truncated 371 aa protein"	
FT		replace(1090,CTTG)	
FT		/*tag= c	"4-nucleotide insertion, results in frameshift and truncated protein"
FT	mutation	replace(1094,A)	
FT		/*tag= h	replace(1284,A)
FT	mutation	/*tag= d	"1-nucleotide insertion, results in frameshift and truncated 422 aa protein"
FT	mutation	/note= "1-nucleotide insertion, results in frameshift and truncated 422 aa protein"	
FT		/*tag= h	"1-nucleotide deletion"
FT	mutation	/note= "1-nucleotide deletion"	
FT		1313	
FT		/*tag= i	"1-nucleotide deletion, results in frameshift and truncated 478 aa protein"
FT	variation	replace(1317,T)	
FT		/*tag= g	"silent polymorphism"
FT	mutation	/note= "silent polymorphism"	
FT		replace(1365,C)	
FT		/*tag= e	"1-nucleotide insertion"
FT	variation	/note= "1-nucleotide insertion"	
FT		replace(1698,T)	
FT		/*tag= r	"silent polymorphism"
FT	misc_feature	/note= "silent polymorphism"	
FT		1995	
FT		/*tag= s	"Alu sequence overlapping with PFKL promoter starts at this position"
FT	polyA_signal	2191..2195	
FT		/*tag= t	
XX			
XX			
XX	W09918197-A2.		
XX			
XX			
XX	15-APR-1999.		
XX			
XX	02-OCT-1998;	98WO-EP006294.	
XX			
XX	02-OCT-1997;	97EP-00117154.	
XX	08-OCT-1997;	97EP-00117398.	
XX	12-NOV-1997;	97EP-00119810.	
XX			
XX			
XX	(PLAC ) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.		
XX	(NAPU-) NAT PUBLIC HEALTH INST.		
XX			
XX			
XX	Peltonen L, Aaltonen J, Björnses P, Perheentupa J, Palotie A, Horelli-Kuitunen N, Yaejo M, Lehtach H;		
XX			
XX	WPI, 1999-287735/24.		
XX			
XX	P-PSDB: AAY06073.		
XX			

XX New polypeptide which co-segregates in mutated form.  
 PS Claim 1, Fig 2a; 77pp; English.  
 XX  
 CC This is the nucleotide sequence of human autoimmune polyglanular disease  
 CC type 1 (APGD1 or AIRE) cDNA. It is a composite of 2 clones, which were  
 CC obtained from a human adult thymus cDNA library by PCR (see AA158609-10),  
 CC and of a 3'UTR extension PCR product. The encoded protein (see AA706073)  
 CC is a transcription factor or transcription-associated factor that may  
 CC associate with vimentin fibres, perhaps as part of a docking mechanism  
 CC regulating nuclear translocation. Aggregates of the mutated protein may  
 CC prevent formation of vimentin intermediate filaments. Mutated APGD1  
 CC polypeptides co-segregate with autoimmune polyendocrinopathy candidiasis  
 CC ectodermal dystrophy (APECED). The invention provides vectors and host  
 CC cells for preparation of APGD1 polypeptides. APECED-associated mutations  
 CC found in the APGD1 gene include: an insertion of 4 nucleotides (CCTG),  
 CC normally found at position 1086-1089, at nucleotide position 1085 or 1090  
 CC ; an insertion of an adenosine at position 1284; an insertion of a  
 CC cytosine at position 1365; a deletion of nucleotides 1085-1097; a  
 CC deletion of the thymidine at position 1051; a deletion of the cytosine at  
 CC position 1309 or 1313; a cytosine to thymidine exchange at position 889;  
 CC a guanosine to thymidine exchange at position 358; an adenosine to  
 CC guanosine exchange at position 374; a guanosine to adenosine exchange at  
 CC position 1052; and a cytosine to adenosine exchange to position 1094. A  
 CC claimed method for testing for carriage of APECED or for the disease  
 CC state involves testing for a mutation in the APGD1 gene or for a mutated  
 CC form of the APGD1 polypeptide. APGD1 nucleic acid can be used in gene  
 CC therapy  
 CC  
 CC Sequence 2245 BP; 419 A; 765 C; 697 G; 364 T; 0 U; 0 Other;

Query Match 66.3%; Score 1023.6; DB 2; Length 2245;

Best Local Similarity	99.08;	Fied. NO. 1.1e-200;
Matches 1026;	Conservative	0;
	Mismatches	4;
	Indels	0;
	Gaps	0;

QY 516 CTGTGCCAGAGAAATGAGACGAGTGTGCTGCTGTCGGACCGCGGGAGCTCATCTCG 575

Db 991 CTCACCCAGAGAAATGAGACGAGTGTGCTGCTGTCGGACCGCGGGAGCTCATCTCG 105

QY	576	TGTGACGGGCTGACCTCGGGGCTTTCACCTGCGCCGTCGTCCTCCCTCCGCTCCGGAGATC	635
Db	1051	TGTGACGGGCTGACCTCGGGGCTTTCACCTGCGCCGTCGTCCTCCCTCCGCTCCGGAGATC	1111

Q7	636	CCCAAGTGGGACCTGGAGGAGTCTCCAGCTGCCTGCAAGGCAACATGTCAGAGAGGTGACAGCCC	695
Db	1111	CCCAAGTGGACCTGGAGGAGTCTCCAGCTGCCTGCAAGGCAACATGTCAGAGAGGTGACAGCCC	117

Db	Sequence	Length
gy	CGGGCAGAGGAGCCCCGGGCCCCCAGAGCACCCTGGAGACCCTGCTCCCCCGGGGCTT	755
1171	CGGGCAGAGGAGCCCCGGGCCCCCAGAGCACCCTGGAGACCCTGCTCCCCCGGGGCTT	123

QY 756 AGGTCGGCGGAGAGAGGTTAAGAGTCCACTTGGGAAACCCCTAAGCCGGCATGACACG 815  
 Db 1221 AGGTCGGCGGAGAGAGGTTAAGAGTCCACTTGGGAAACCCCTAAGCCGGCATGACACG 129

Db	Seq	Score
Qy	816 ACCTCTGTCTACAAGACCTGCGGGCTCGGCCTTGTGACGCCCGCTGCAGGTGTGAC	875
Db	1231 ACCTTGTCTACAAGCACTGCGGGCTCGGCCTTGTGACGCCCGCTGCAGGGCTGAC	135

07 876 TCTTCGGGCTGCACCCCTACTGCTGTGTGGTCTCTGAGGGTCAGACGAACCTGGCTCT 935  
Db 1351 TCTCGGCGCTGCACCCCTACTGCTGTGTGGTCTCTGAGGGTCAGACGAACCTGGCTCT 1410

958 GGTGGCCGTTGCGGGGTGTGCGGAGATGGTACCGACGTGCGGTGTACTCACTACGCGCC 147  
 1411 GGTGGCCGTTGCGGGGTGTGCGGAGATGGTACCGACGTGCGGTGTACTCACTACGCGCC 147

1471 GCTGCTTCCACTGGCGCTGCCACTTCCAGCGCGACACTCCCGCGCCGGGACGGGACTG 153

CCGCGAGACCTCCACGAGAGCAGTCCCCAGCCCCGGTGGAGGGGGGTCCGCCCCC

1058





Db 1587 TCACGGCATCTGCAGTGGGCCATCCAGCATGTTCACGCCCGCTGGCCGAGACACACC 1646  
QY 1280 CT 1281  
Db 1647 CT 1648

RESULT 6  
ABL32092  
ID ABL32092 standard; DNA; 9876 BP.

AC ABL32092;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 65.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antihaemic; cyrostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PP 02-JUL-2001; 2001WO-EP007537.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.

PS Claim 1; SEQ ID NO 65; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention

XX Sequence 9876 BP; 1912 A; 309 C; 2928 G; 4727 T; 0 U; 0 Other;

SQ Query Match 7.7%; Score 118.4; DB 6; Length 9876;

Best Local Similarity 76.0%; Pred. No. 1.6e-15;

Matches 146; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 AGAGAAAGTGAAGTCTTCTCAGGCTTTAAGACATGGCGTTTGTCCAGGCTGAACCG 60

Db 9386 AAGAGAAAGTGAAGTCTTCTCAGGCTTTAAGAGTATGGCGTTTGTGATATCG 9445

QY 61 CTGCTCTCAGCTGGCGCCGCTGGGTGGCGCGGCCCTGCTATACCGAGAGTCAAG 120

Db 9446 TTGTTTATTAGTGGTTCGTGGGTGGGTGGCGCGGTTTGTATTAGTGAAGATTAAAG 9505

QY 121 ATCCATGGGAATGCATGCTCATCTTCGTCGCCAGCATGTTCTTAATGGGTAAGA 180  
Db 9506 ATTATTTGGAATGTTATGTTATTTTCGTTTAAATGATGTTTAAATGAGGATAGA 9565  
QY 181 GCAGGTGGGAG 192  
Db 9566 GTAGTGTGGCGG 9577

RESULT 7  
ABL32093/C  
ID ABL32093 standard; DNA; 9876 BP.

AC ABL32093;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 66.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antihaemic; cyrostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PP 02-JUL-2001; 2001WO-EP007537.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.

PS Claim 1; SEQ ID NO 66; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention

XX Sequence 9876 BP; 1882 A; 309 C; 3154 G; 4531 T; 0 U; 0 Other;

SQ Query Match 6.6%; Score 101.8; DB 6; Length 9876;

Best Local Similarity 69.2%; Pred. No. 5.6e-12;

Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 AGAGAAAGTGAAGTCTTCTCAGGCTTTAAGACATGGCGTTTGTCCAGGCTGAACCG 60

Db 491 AAAAAAATAAATCTTCTCAAACTTAAACATAGCTTAATCCAACTATACCG 432

QY 61 CTGCTCTCAGCTGGCGCCGCTGGGTGGCGCGGCCCTGCTATACCGAGAGTCAAG 120

Db 431 CTACTCTCAACTAAACCGTAAATTAACCGAACGCCCTTAATTAACCAAAAAATCAAAA 372

QY 121 ATCCACTGGGAATGCCATCTCATCTTTGTCCTCCGACATGGTTCTTAATGGGTAGAA 180  
DB 371 ATCCACTGAATAAATACATCTATCTTGTGCTCCCAACATTAATCTTAATAAATAA 312  
QY 181 GCAGGTCGGAGAGAACCTGCC 201  
DB 311 ACATATATATAAATAATCTTACC 291

## RESULT 8

ADH12179  
ID ADH12179 standard; cDNA; 9646 BP.

XX AC ADH12179;

XX DT 11-MAR-2004 (first entry)

DE Human CHD5 (chromodomain, helicase, DNA-binding 5)-encoding cDNA.

XX Human; chromodomain helicase DNA-binding 5; CHD5; chromosome 1p36.3;  
XX chromatin structure; chromatin unwinding; DNA repair; DNA recombination;  
XX transcriptional regulation; gene expression; cell cycle control;  
XX development regulation; oncogenesis; brain; neural development;  
XX neural tissue neoplasia; diagnosis; cancer; neural cancer; neuroblastoma;  
XX breast cancer; colon cancer; liver tumour; germ cell tumour;  
XX drug screening; cytostatic; gene therapy; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 101..5965

FT FT /tag=a

FT /product= "CHD5"

XX WO2003106650-A2.

XX 24-DEC-2003.

XX 16-JUN-2003; 2003WO-US019027.

XX 14-JUN-2002; 2002US-0388848P.

XX (CHIL-) CHILDBRENS HOSPITAL PHILADELPHIA.

XX Brodeur GM, White PS;

XX WPI: 2004-082186/08.

XX P-PSDB; ADH12180.

XX Novel chromodomain helicase DNA-binding (CHD) proteins, useful as  
XX diagnostic and prognostic indicator of tumor, comprises amino terminus  
XX having two PHD class zinc finger domains and two chromodomains.

XX Claim 5; SEQ ID NO 1; 124pp; English.

XX The invention relates to human chromodomain, helicase, DNA-binding 5  
XX (CHD5; ADH12180) and cDNA encoding it (ADH12179). CHD5 is a novel member  
XX of the CHD gene family, members of which have a profound effect on  
XX chromatin structure and gene expression and which are thus likely to play  
XX an important role in cell cycle control, regulation of development, and  
XX oncogenesis. CHD5 comprises two N-terminal zinc finger domains of the PHD  
XX (plant homeodomain) class, two chromodomains, a central region which  
XX contains a predicted DEAH-box-type helicase domain and a putative SNF2  
XX domain, and several nuclear localization signals. The gene encoding CHD5  
XX is located on chromosome 1p36.3, a region frequently deleted in a variety  
XX of cancers including neuroblastoma, and the protein is preferentially  
XX expressed in brain. CHD5 is therefore thought to be a modulator of normal  
XX neural development and neoplasias of neural tissue origin. The invention  
XX also relates to vectors and host cells comprising the CHD5 cDNA sequence;  
XX an antibody against CHD5; a method of screening for modulators of CHD5  
XX activity; a method of diagnosing cancer in a patient, where a reduced  
XX level or absence of CHD5 or CHD5 nucleic acids indicates the presence of

CC cancer; treating cancer by administration of CHD5 protein, CHD5-encoding  
CC nucleic acids or CHD5 mimetics; and CHD5-specific PCR primers (ADH12186-  
CC ADH12247). The methods of the invention are useful in the diagnosis or  
CC treatment of cancers such as neural cancers (e.g., neuroblastoma), breast  
CC cancer, colon cancer, liver tumours and germ cell tumours. The CHD5  
CC protein, CHD5 nucleic acids and anti-CHD5 antibodies are useful as  
CC research tools to identify other proteins that are intimately involved in  
CC chromatin unwinding, DNA repair and recombination, and transcriptional  
CC regulation. The present sequence represents cDNA encoding human CHD5.  
XX

XX Sequence 9646 BP; 2076 A; 2789 C; 3013 G; 1768 T; 0 U; 0 Other;

Query Match 4.2%; Score 65; DB 12; Length 9646;

Best Local Similarity 62.7%; Pred. No. 0.0004;  
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGACGAGTGTCCCGTGTGGGACGGGGGAGCTCATCTGTGACGGGCTCC 591

DB 1128 AGGATTACTGTAGTGTGCTCCAGCAGGCTGGGAGATCTCTGTCCGACACTGCCCA 1187

QY 592 GGGCTTCACCTGTGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGGACCTGA 651

DB 1188 GGGCTTACCATCTCGTATGCTGACCCAGACTGGAGAAAGGCTCCGAGGCAAGTGA 1247

QY 652 GGTGCTCCAGTCTGCTGCGCAACAGTCCAGAGGTGAG 692

DB 1248 GCTGCCCCCATCTGTGAGAGAGGGGATCCAGTGGAGCCG 1288

## RESULT 9

ABN38615  
ID ABN38615 standard; DNA; 60 BP.

XX AC ABN38615;

XX DT 15-JUL-2002 (first entry)

DE Human spliced transcript detection oligonucleotide SEQ ID NO:11363.

XX Human; mouse; rat; splice transcript; detection; RNA transcript;

XX splice variant; transcriptome; oligonucleotide library; ss.

XX OS Homo sapiens.

XX WO200210449-A2.

XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-1B001903.

XX 28-JUL-2000; 2000US-0221607P.

XX 02-MAY-2001; 2001US-0287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI: 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which  
XX selectively hybridize to mRNAs transcribed from a transcript unit of a  
XX genome, useful for detecting tissue-, pathology-, and developmental-  
XX specific genes.

XX Example 1; SEQ ID NO 11363; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
XX messenger RNAs that populate a (sub-)transcriptome, where the (sub-  
XX )transcriptome comprises messenger RNAs transcribed from multiple  
XX transcription units that populate a genome. The library comprises several  
XX oligonucleotides, each capable of hybridizing selectively to a set of  
XX messenger RNAs transcribed from a given transcription unit of the genome,  
XX which encodes one or more messenger RNA splice variants. The

oligonucleotide libraries are useful for detecting mRNAs from a biological sample. In expression profiling studies, in qualitatively or quantitatively characterizing the corresponding transcriptome, and in detecting RNA transcripts and splice variants of human or animal transcriptomes. The libraries may also be used as specialised mini libraries to detect transcripts of a sub-transcriptome under a particular biological or pathological state, and so allowing the detection of tissue - and pathology-specific genes such as those genes only expressed in specific tissue under a specific pathological condition; to detect developmental specific genes; and to detect RNA transcripts and splice variants of a transcriptome of a patient suffering from a particular disorder. ABN7253 to ABN59589 represent oligonucleotide sequences from rats, humans and mice, which are used in the exemplification of the present invention. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://www.intellectualproperty.wipo.int/pub/published\\_pct\\_sequences](http://www.intellectualproperty.wipo.int/pub/published_pct_sequences)

50 Sequence 60 BP; 13 A; 15 C; 16 G; 16 T; 0 U; 0 Other;  
Query Match 3.9%; Score 60; DB 6; Length 60;  
Best Local Similarity 100.0%; Pred. No. 0.0023;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1403 GCTGGGCTGTAACAGCTCTGTGTTTCTGGGGACACAGCCATCATATGTGCTGGAATTA 1462

Db 1 GCTGGGCTGTAACAGCTCTGTGTTTCTGGGGACACAGCCATCATATGTGCTGGAATTA 60

RESULT 10  
ABK70271  
ID ABK70271 standard; cDNA; 5739 BP.

DT 15-JUL-2002 (first entry)

DE Human lung cancer associated full length cDNA DMSM-6.

KW Human; ss; gene; lung cancer; cytostatic; tumour; vaccine.

OS Homo sapiens.

PN W0200224057-A2.

PD 28-MAR-2002.

PF 20-SEP-2001; 2001WO-US042232

PR 22-SEP-2000; 2000US-0234837P.

PR 29-JUN-2001; 2001US-0301928P.

PA (CORI-) CORIXA CORP.  
PA  
CORI-

PI Benson DR, Mohamath

DR WPI; 2002-372001/40.

PT New tumor lung prote

PT cancer.

PS Claim 1; Page 146-148; 189pp; English.

The invention relates to an isolated polynucleotide comprising a sequence selected from 183 human DNA sequences (appearing as ABR70130-ABR70312) or their fragments, homologues, variants or complements and their encoded polypeptides. Also included are an expression vector comprising the polynucleotide operably linked to an expression control sequence; a host cell transformed or transfected with an expression vector of, an isolated antibody, or its antigen-binding fragment that specifically binds to the polypeptide; a method for detecting the presence of a cancer in a patient; a fusion protein comprising at least the polypeptide; an

oligonucleotide that hybridises to the polynucleotide under moderately stringent conditions; a method for stimulating and/or expanding T cells specific for a tumour protein; an isolated T cell population comprising T cells prepared from the method of above; a composition comprising a first component consisting of carriers and immunostimulants, and a second component selected from the polynucleotides, proteins, antibodies, fusion proteins, T cell populations and antigen presenting cells expressing the polynucleotide; methods for stimulating an immune response or treating cancer in a patient by administering the composition and diagnostic kits comprising at least one of the oligonucleotide of, or an antibody and a detection reagent consisting of a reporter group. The polypeptides and polynucleotides are useful as vaccines for the treatment or prevention of lung cancer, and for diagnosis and monitoring of such cancer. The polynucleotide, polypeptide and antigen presenting cells can be used to stimulate or expand T cells specific for a tumorous protein. The polynucleotides may be used as probes or primers for nucleic acid hybridisation, and in the preparation of ribozyme molecules for inhibiting expression of tumour polypeptides and proteins in tumour cells. The present sequence is one of the 183 lung cancer associated polynucleotides

Query Match	3.9%	Score 60;	DB 6;	Length 5739;
Best local Similarity	61.5%;	Pred. No. 0.0043;		
Matches 96;	Conservative 0;	Mismatches 60;	Indels 0;	Gaps 0;
SQ Sequence 5739 BP; 1618 A; 1323 C; 1601 G; 1197 T; 0 U; 0 Other;				

QY 532 AGGACGAGTGTGCGGTCGTGTCGGGACGGCGGGGAGCTATCTGCTGTACGGGTCGCCCTTC 591

Db 1109 AGGCTATATGCGAGGTGTGCTCCAGGCAAGGCGGTGATCATCTCTGTGTATATACCTGTCCCC 1168

QY 592 GGAGCTTCCACCTGGCCTGCTGCCCTCCGCTCGGAGATCCCACTGGACCTGGA 651

Db 1169 GTGGTACCAATGGTCTGCTGATCCGACATGAGAGGCTCCCGAGGCGAAGTGA 1228

QY 652 GGTGCTCCAGCTGCTGCAGGCAACAGTCCAGAGG 687  
| | | | | | | | | | | | | | | | | |  
Db 1229 GCTGCCACACTGTCGAGAAGGACATCCAGTGGG 1264

RESULT 11  
AAT32301  
ID AAT32301 standard: cDNA: 6327 BP.

30-OCT-1996 (first entry)

DE Dermatomyositis specific autoantigen, Mi-2, coding sequence.

Mi-2: autoantigen: collagen disease: chromosome 12; 12p13; helicase;

[illegible]

XX 8

ET

ET :

ET :

ET :

ET :

PN

PD

PF

PR

XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.  
 XX Seelig HP, Renz M;  
 XX WPI: 1996-240280/25.  
 DR P-PSDB; AAR99334.  
 XX  
 PT DNA encoding dermatomyositis specific auto:antigen - useful for  
 PT differential diagnosis and treatment of dermatomyositis.  
 XX  
 PS Claim 3: Fig 2: 20pp; German.  
 XX  
 CC The present sequence encodes a 218 kD dermatomyositis specific auto-  
 CC antigen, designated M1-2. The sequence numbering given in the  
 CC specification starts at nucleotide 91, i.e. the 5' UTR is omitted, as  
 CC well as the 'A' of the first ATG start codon. The protein is hydrophilic,  
 CC acidic and protruding regions of the protein are characteristic of  
 CC helices. The gene corresponding to this cDNA was localised to  
 CC chromosome 12 (12p13). The DNA can be used for the recombinant production  
 CC of M1-2 which is used for, e.g. the differential diagnosis of collagen  
 CC diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting  
 XX  
 SQ Sequence 6327 BP; 1761 A; 1473 C; 1740 G; 1353 T; 0 U; 0 Other;  
 Query Match 3.9%; Score 60; DB 2; Length 6327;  
 Best Local Similarity 61.5%; Pred. No. 0.0044;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 532 AGGACGAGTGTCCCGTGTCTGGGAGCGGGAGCTCATCTGCTGAGCGCTGCCCTC 591  
 DB 1108 AGGACGAGTGTCCCGTGTCTGGGAGCGGGAGCTCATCTGCTGAGCGCTGCCCTC 1167  
 QY 592 GGGCCCTTCACCGTGGCTGCTGCTCCCTCCGCTCCGGGAGATCCCGAGTGGAGCTGGA 651  
 DB 1168 GTGCTTACCAATGCTGCTGCTGCTGATCCGACATGAGAGAGCTCCCGAGGAGCAATGGA 1227  
 QY 652 GGTGCTCCAGCTGCTGCTGAGGCAACAGTCCAGAGG 687  
 DB 1228 GCTGCCACATGCTGCGAGAGGAGCATCCAGTGGG 1263  
 RESULT 12  
 ABL64410  
 ID ABL64410 standard; DNA; 6417 BP.  
 XX  
 AC ABL64410;  
 XX  
 PS 15-MAY-2002 (first entry)  
 XX  
 DE Stomach cancer related gene sequence SEQ ID NO:2747.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytotoxic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200194629-A2.  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 18-SEP-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233137P.  
 PR 20-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 PA (AVAL-) AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 DR WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1: SEQ ID NO 2747; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (II) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' s  
 CC tumour  
 XX  
 SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;  
 Query Match 3.9%; Score 60; DB 6; Length 6417;  
 Best Local Similarity 61.5%; Pred. No. 0.0044;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 532 AGGACGAGTGTCCCGTGTCTGGGAGCGGGAGCTCATCTGCTGAGCGCTGCCCTC 591

Db 1198 AGAGCTATTGCCAGGTGTGCAGACAAGGGGTGAATCATCCGTGTGTAACCTGTCCCC 1257  
QY 592 GGGACCTTCCACCTGGCTCTGCTCCCTCCGTCGGGAGATCCCAAGTGGAGACTGTGA 651  
Db 1258 GTGCTTACCAATGATGTCTGCTGTGATCCCGACATGGAAGAGCTCCCGAGGCAAGTGA 1317  
QY 652 GGTGCTCCAGCTGCTGTCAGGCAACAGTCCAGGAGG 687  
Db 1318 GCTGCCACACTGCGAAGGAAGGCAATCCAGTGGG 1353

RESULT 13  
ABX04169  
ID ABX04169 standard; cDNA; 6417 BP.  
AC ABX04169;  
XX  
DT 10-JAN-2003 (first entry)  
DE Human mRNA differentially expressed in mesenchymal cells #16.  
XX  
KM Human; ss; gene; skeletal growth; cartilage degeneration disorder;  
KM chondroblastic phenotype; mesenchymal cell; cartilage formation;  
KM bone formation; arthritis; osteoarthritis; rheumatoid arthritis;  
KM gout arthritis; adjuvant arthritis; arthritis deformans; ankylosis;  
KM infectious arthritis; osteochondrosis; RDA; antihypertic; osteopathic;  
KM antineumatic; antineumatic; representational difference analysis.  
XX  
OS Homo sapiens.  
XX  
PN WO200271927-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 12-MAR-2002; 2002WO-US007787.  
XX  
PR 12-MAR-2001; 2001US-0274980P.  
PA (BGMH) BRIGHAM & WOMENS HOSPITAL INC.  
PI Yates KE, Mizuno S, Glowacki J;  
DR WPI; 2002-723276/78.  
XX  
PT New nucleic acid molecules capable of promoting chondrogenesis, useful  
PT for diagnosing and treating cartilaginous tissue degeneration conditions,  
PT e.g. osteoarthritis, rheumatoid arthritis, gout arthritis, or  
PT osteochondrosis.  
XX  
PS Claim 33; Page 111-113; 153pp; English.  
XX  
CC The invention relates to new isolated nucleic acid molecule comprising a  
CC nucleic acid molecule consisting of a gene differentially expressed in  
CC cells undergoing differentiation from mesenchymal cell to a  
CC chondroblastic phenotype, or hybridising under stringent conditions to  
CC them (or their fragments). Also included are expression vectors,  
CC transformed host cells, expressed polypeptides or peptide fragments  
CC (which induce differentiation of a mesenchymal cell and may be used as an  
CC immunogen), binding partners of the polypeptides, a method for  
CC identifying an agent useful in modulating mesenchymal cell  
CC differentiation induction activity of a molecule, a method of diagnosing  
CC a condition characterized by aberrant expression of a nucleic acid  
CC molecule or its expression product; a method for determining regression,  
CC progression or onset of cartilaginous tissue degeneration condition in a  
CC subject characterised by aberrant expression of a nucleic acid molecule  
CC or its expression product, a method for treating a cartilaginous tissue  
CC degeneration condition, a method for treating a subject to reduce the  
CC risk of cartilaginous tissue degeneration condition developing in the  
CC subject, a method for identifying a candidate agent for treating a  
CC cartilaginous tissue degeneration condition, and a solid-phase nucleic  
CC acid molecule array consisting essentially of a set of nucleic acid  
CC molecule as cited above (or known from known genes shown to be

CC differentially expressed in developing mesenchymal cells using the  
CC technique of representational difference analysis, RDA), its expression  
CC products or fragments, fixed to a solid substrate. The nucleic acids,  
CC polypeptides and agents are useful for treating cartilaginous tissue  
CC degeneration conditions such as osteoarthritis, rheumatoid arthritis,  
CC gout arthritis, adjuvant arthritis, arthritis deformans, infectious  
CC arthritis or osteochondrosis. The present sequence is a cDNA from a  
CC known gene differentially expressed in developing mesenchymal cells  
XX  
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;  
Query Match 3.9%; Score 60; DB 6; Length 6417;  
Best Local Similarity 61.5%; Pred. No. 0.0044;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 532 AGAGCAAGTGTGCTGCTGTGTCAGACGGGGGAGACTCATCTGCTGTGACGGCTGCTCC 591  
Db 1198 AGAGCTATTGCCAGGTGTGCAGACAAGGGGTGAATCATCCGTGTGTAACCTGTCCCC 1257  
QY 592 GGGACCTTCCACCTGGCTCTGCTCCCTCCGTCGGGAGATCCCAAGTGGAGACTGTGA 651  
Db 1258 GTGCTTACCAATGATGTCTGCTGTGATCCCGACATGGAAGAGCTCCCGAGGCAAGTGA 1317  
QY 652 GGTGCTCCAGCTGCTGTCAGGCAACAGTCCAGGAGG 687  
Db 1318 GCTGCCACACTGCGAAGGAAGGCAATCCAGTGGG 1353

RESULT 14  
ABR84696  
ID ABR84696 standard; cDNA; 6417 BP.  
XX  
AC ABR84696;  
XX  
DT 14-AUG-2002 (first entry)  
DE Human cDNA differentially expressed in granulocytic cells #1267.  
XX  
KM Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KM viral infection; parasitic infection; protozoal infection;  
KM fungal infection; sterile inflammatory disease; psoriasis;  
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KM cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KM adult respiratory distress syndrome; inflammatory bowel disease;  
KM Crohn's disease; ulcerative colitis; periodontal disease;  
KM granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US030821.  
XX  
PR 03-OCT-2000; 2000US-0237189P.  
PA (GENE-) GENE LOGIC INC.  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
DR WPI; 2002-435328/46.  
XX  
PT Detecting granulocyte activation by detecting differential expression of  
PT genes associated with granulocyte activation, which serves as diagnostic  
PT markers that is useful for monitoring disease states and drug toxicity.  
XX  
PS Claim 1; SEQ ID NO 1267; 114pp; English.  
XX  
CC The invention relates to detecting (M) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing the  
CC expression level to an expression level in an unactivated GC, where









```
XX ss; cancer; neurodegenerative disease; human.
KM Homo sapiens.
OS US2003104529-A1.
PN 05-JUN-2003.
PD 04-JAN-2002; 2002US-00037270.
XX 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX 19-JUL-2000; 2000US-00620312.
XX (ZHOU/) ZHOU P.
PA (TANG/) TANG Y T.
PA (LIU/) LIU C.
PA (ASUNDI/) ASUNDI V.
PA (DRMANAC/) DRMANAC R T.
XX Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX WPI; 2003-678194/64.
XX New polynucleotide, useful for treating diseases e.g., cancer or
PT neurodegenerative diseases.
PS Claim 1; SEQ ID NO 325; 99pp; English.
XX The invention relates to a polynucleotide comprising a sequence given in
CC the specification, or its mature protein-coding portion, or its
CC complement. The polynucleotide is useful for treating diseases e.g.,
CC cancer or neurodegenerative diseases and many others listed in the
CC specification. The present sequence represents a novel human cDNA. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=2003104529.
XX SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;
Query Match 3.9%; Score 60; DB 9; Length 6475;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCCTGTGTGTCGGGAGCGGGGAGTCTATCTGCTGTGACGGTCCCTC 591
DB 1327 AGGACTATTGCGAGGTGTGTCGAGCAAGGCGGTGAGATCTCTGTGTATCTCTGTCC 1386
QY 592 GGGCCTTCCACCTGGGCTGCTGTCCTCCCTCCGCTCCGGGAGATCCCGAGTGGGAGCTTGA 651
DB 1387 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAAAGGCTCCCGAGGCAATGGA 1446
QY 652 GGTGCTCCAGCTGCTGTCGAGGCAAGCATGTCAGAGG 687
DB 1447 GCTGCCACACTGCGAGAAAGGATCCATCGTGG 1482
RESULT 21
ID AAI60230 standard; cDNA; 6553 BP.
XX AAI60230;
AC 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4219.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
```

```
KM Leukaemia; ss.
XX Homo sapiens.
OS WO200153312-A1.
XX 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX P-PSDB; AAM41074.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
PS Claim 1; SEQ ID NO 4219; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression.
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX SQ Sequence 6553 BP; 1802 A; 1538 C; 1829 G; 1384 T; 0 U; 0 Other;
Query Match 3.9%; Score 60; DB 4; Length 6553;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCCTGTGTGTCGGGAGCGGGGAGTCTATCTGCTGTGACGGTCCCTC 591
DB 1304 AGGACTATTGCGAGGTGTGTCGAGCAAGGCGGTGAGATCTCTGTGTATCTCTGTCC 1363
QY 592 GGGCCTTCCACCTGGGCTGCTGTCCTCCCTCCGCTCCGGGAGATCCCGAGTGGGAGCTTGA 651
DB 1364 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAAAGGCTCCCGAGGCAATGGA 1423
QY 652 GGTGCTCCAGCTGCTGTCGAGGCAAGCATGTCAGAGG 687
DB 1424 GCTGCCACACTGCGAGAAAGGATCCATCGTGG 1459
RESULT 22
ID ADQ22446 standard; DNA; 6848 BP.
XX ADQ22446
```



KW kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;

KM apoptotic disorder; cell proliferation disorder.  
 XX Homo sapiens.  
 OS  
 PN WO200299122-A1.  
 XX  
 PD 12-DEC-2002.  
 XX  
 PF 03-JUN-2002; 2002MO-US017382.  
 XX  
 PR 05-JUN-2001; 2001US-0296076P.  
 PR 10-OCT-2001; 2001US-0328605P.  
 PR 15-FEB-2002; 2002US-0357253P.  
 XX  
 PA (EXEL-) EXELIXIS INC.  
 PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP,  
 XX  
 DR WPI: 2003-156859/15.  
 DR P-PSDB; AB007199.  
 XX  
 PT Identifying modulators of the p53 pathway for use in treating apoptotic  
 PT or cell proliferation disorders, comprises screening for agents that  
 PT modulate activity of a human ortholog of genes that modify the p53  
 PT pathway in *Drosophila*.  
 XX  
 PS Example 2; Page 269-271; 678pp; English.  
 XX  
 CC The invention relates to identifying (M1) a candidate p53 pathway  
 CC modulating agent, by contacting an assay system comprising a purified HM  
 CC polypeptide (human orthologue of genes that modify the p53 pathway in  
 CC *Drosophila*) or nucleic acid with a test agent under conditions, where but  
 CC for the presence of the test agent, the system provides a reference  
 CC activity, and detecting a test agent-biased activity of the assay system.  
 CC Also included are modulating (M2) a p53 pathway of a cell (comprising  
 CC contacting a cell defective in p53 function with a candidate modulator  
 CC that specifically binds to a HM polypeptide comprising an HM amino acid  
 CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
 CC in a mammalian cell (comprising contacting the cell with an agent that  
 CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
 CC a disease in a patient (comprising: (a) obtaining a biological sample  
 CC from the patient; (b) contacting the sample with a probe for HM  
 CC expression; (c) comparing the results with a control; and (d) determining  
 CC whether the comparison indicates a likelihood disease). (M1) is useful  
 CC for identifying modulators of the p53 pathway. A probe for HM expression  
 CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
 CC in a patient, where the cancer has greater than 25 % expression level.  
 CC Modulators identified by (M1) are useful in a variety of diagnostic and  
 CC therapeutic applications, where disease or disorder prognosis is related  
 CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
 CC proliferation disorders (e.g., cancer). Another two new methods (M2 and  
 CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
 CC the p53 function of the cell, so that the cell undergoes normal  
 CC proliferation or progression through the cell cycle. (M2) and (M3) are  
 CC also useful for treating defects in the p53 pathway such as angiogenic,  
 CC apoptotic or cell proliferation disorders. The present sequence is an HM  
 CC nucleic acid encoding a p53 pathway modifying protein  
 XX  
 SQ Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;  
 XX  
 Query Match 3.9%; Score 59.8; DB 8; Length 6331;  
 Best Local Similarity 58.9%; Pred. No. 0.0046; Mismatches 72; Indels 0; Gaps 0;  
 Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
 QY 532 AGGACAGTGTGCGGTGTGCGGAGCGGAGGCTCATCTGTGACGCGTGCCTC 591  
 DB 1346 AGGATTACTGTGAGGTGTGCGACAGGCTGCGGAATTAATCTGTGACACTGCGCTC 1405  
 QY 592 GGGCCCTTCACTGCGCTGCTTCCCTCCGTCGCGGAGATCCCAAGTGGACCTGGA 651  
 DB 1406 GTGCGCTTACACCTGCTGTGCTTGAATCTGAGCTTACCGGCGCTCCAGAGGCGCAATGGA 1465  
 QY 652 GGTGCTCCAGCTGCTGCGAGGCAAGTCCAGAGGTGACGCCCGGCGAGAGGA 706

Db 1466 GCTGCCCTCACTGTGAGGAGGCGGCTCCAGTGGGAGGCCAAGAGAGAGAGA 1520  
 RESULT 27  
 AD126769  
 ID AD126769 standard; DNA; 6331 BP.  
 XX  
 AC AD126769;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Human CHD3 DNA sequence, SEQ ID 30.  
 XX  
 KW Cytostatic; ophthalmological; Antiarthritic; Antirheumatic;  
 KW Antiinflammatory; Antiarteriosclerotic; Antipsoriatic; Gynaecological;  
 KW Vasotropic; Vulnerary; Hypotensive; Antiallginal; Cardiant; Osteopathic;  
 KW Gene therapy; antiangiogenic; tumour vascularisation; retinopathy;  
 KW rheumatoid arthritis; Crohn disease; atherosclerosis;  
 KW ovarian hyperstimulation; psoriasis; endometriosis; neovascularization;  
 KW restenosis; cardiovascular disease; hypertension; Raynaud disease;  
 KW muscular degeneration; osteoporosis; human; CHD3; gene; de.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 211..6045  
 FT /\*tag= a  
 FT /\*product= "CHD3"  
 XX  
 PN FR837391-A1.  
 XX  
 PD 26-SEP-2003.  
 XX  
 PF 22-MAR-2002; 2002PR-00003655.  
 XX  
 PR 22-MAR-2002; 2002PR-00003655.  
 XX  
 PA (GENE-) GENE SIGNAL.  
 PA (ALMA/) AL MAHMOOD S.  
 XX  
 PI Colin S. Schneider C, Al Mahmood S;  
 XX  
 DR WPI: 2003-806304/76.  
 DR P-PSDB; AD126773.  
 XX  
 PT Antiangiogenic composition, useful for treating e.g. tumors, comprises  
 PT new angiogenesis-related nucleic acids or related polypeptides,  
 PT antibodies or antisense sequences.  
 XX  
 PS Claim 2; SEQ ID NO 30; 110pp; French.  
 XX  
 CC The present invention relates to antiangiogenic pharmaceutical  
 CC compositions. The compositions comprise, as active ingredient, one or  
 CC more of: nucleic acid (I; AD126740-AD126744 and AD126766-AD126769) from  
 CC an endothelial cell gene the expression of which is induced by an  
 CC angiogenic agent, or its complement or fragment; polypeptide (II;  
 CC AD126745-AD126749 and AD126770-AD126773) encoded by (I), or its fragment;  
 CC an antisense nucleic acid (III; AD126750-AD126753) able to inhibit  
 CC expression of (I); or an antibody (Ab) that binds (II). The compositions  
 CC are used for treating a wide range of angiogenic diseases: tumour  
 CC vascularisation; retinopathy; rheumatoid arthritis; Crohn disease;  
 CC atherosclerosis; ovarian hyperstimulation; psoriasis; endometriosis;  
 CC associated with neovascularization; restenosis (arterial or after balloon  
 CC angioplasty); overgrowth of cells in wound healing; peripheral vascular  
 CC disease; hypertension; vascular inflammation; Raynaud disease; aneurysm;  
 CC thrombophilic; lymphedema; lechemia; angina; myocardial infarct;  
 CC chronic heart disease; (congestive) cardiac insufficiency; age-related  
 CC muscular degeneration and osteoporosis. This sequence encodes  
 CC chromodomain helicase DNA binding protein 3 (CHD3).  
 XX  
 SQ Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;

	Query Match	3.9%;	Score 59.8;	DB 10;	Length 6331;	
	Best Local Similarity	58.9%;	Pred. No. 0.0046;			
	Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0	
Oy	5312 AGGAGAGAGTGTCCCGTGATGTCTGGGACGCGGGGAGACTCAATCTGTGACGAGTCGCCCTC	591				
Dd	1346 AGGATTACTGTAGAGTGTCGCCAGCAGGGGTGGGGAAATTATTCTGTGTGACACTGCCTC	140				
Oy	552 GGGCTTTCCACTGCGCTGCTGCTGCTCCCTTCCTCCGCGAAGATCCCAGTGGGACTCTGA	651				
Dd	1406 GGCTCTACCACCTCGCTGCTGCTTGAATCTGAGCTTGACCGGGCTCCAGAGGGCAAATVGA	146				
Oy	652 GGTGCTCCAGCTGCGCTCGCAGGCAACGATCCAGAGGTGACGCCCGCGGACAGAGA	706				
Dd	1466 GCTGCTCATCTGTGTGAGAAAGAGGGGGTTCAGTGGAGGCCCAAGAGGAAGAAAGA	1520				
RESULT 28						
ID	AD126768					
XX	AD126768 standard; DNA; 6331 BP.					
AC	AD126768;					
XX						
DT	22-APR-2004 (first entry)					
DE	Human CHD3 DNA sequence, SEQ ID 29.					
XX						
KW	Cytostatic; ophthalmological; Antiarthritic; Antirheumatic;					
KW	Antiinflammatory; Antiarteriosclerotic; Antipsoriatic; Gynaecological;					
KW	Vasectopic; Vulnereary; Hypotensive; Antianginal; Cardiant; Osteopathic;					
KW	Gene therapy; angiogenic; tumour vascularisation; retinopathy;					
KW	rheumatoid arthritis; Crohn disease; atherosclerosis;					
KW	ovarian hyperstimulation; psoriasis; endometriosis; neovascularization;					
KW	resenosis; cardiovascular disease; hypertension; Raynaud disease;					
KW	muscular degeneration; osteoporosis; human; CHD3; gene; ds.					
OS	Homo sapiens.					
XX						
FH	Key Location/Qualifiers					
FT	CDS 211..6045					
FT	/tag= a					
FT	/product= "CHD3"					
XX						
FN	FR2837391-A1.					
XX						
PD	26-SEP-2003.					
XX						
PE	22-MAR-2002; 2002FRR-00003655.					
XX						
PR	22-MAR-2002; 2002FRR-00003655.					
XX						
PA	(GENE-) GENE SIGNAL.					
PA	(ALMA/) AL MAHMOOD S.					
XX						
PI	Colin S, Schneider C, Al Mahmood S;					
DR	WPI; 2003-806304/76.					
XX						
DR	P-PSDB; AD126772.					
XX						
PT	Antiangiogenic composition, useful for treating e.g. tumors, comprises					
PT	new angiogenesis-related nucleic acids or related polypeptides,					
PT	antibodies or antisense sequences.					
PS	Claim 2; SEQ ID NO 29; 110P; French.					
XX						
CC	The present invention relates to antiangiogenic pharmaceutical					
CC	compositions. The compositions comprise, as active ingredient, one or					
CC	more of: nucleic acid (I; AD126740-AD126744 and AD126766-AD126769) from					
CC	an endothelial cell gene the expression of which is induced by an					
CC	angiogenic agent, or its complement or fragment; polypeptide (II;					
CC	AD126745-AD126749 and AD126770-AD126773) encoded by (I), or its fragment;					
CC	an antisense nucleic acid (III; AD126750-AD126753) able to inhibit					
CC	expression of (I); or an antibody (Ab) that binds (II). The compositions					

CC	are used for treating a wide range of angiogenic diseases: tumour
CC	vascularisation; retinopathy; rheumatoid arthritis; Crohn disease;
CC	atherosclerosis; ovarian hyperstimulation; psoriasis; endometriosis;
CC	associated with neovascularisation; restenosis (arterial or after balloon
CC	angioplasty); overgrowth of cells in wound healing; peripheral vascular
CC	disease; hypertension; vascular inflammation; Raynaud disease; aneurysm;
CC	thrombophilicities; lymphedema; ischaemia; angina; myocardial infarct;
CC	chronic heart disease; (congestive) cardiac insufficiency; age-related
CC	muscular degeneration and osteoporosis. This sequence encodes CHD3.
XX	
SO	Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;
XX	
Query Match	3.9%; Score 59.8; DB 10; Length 6331;
Best Local Similarity	58.9%; Pred. No. 0.0048;
Matches 103; Conservative	0; Mismatches 72; Indels 0; Gaps 0;
Oy	532 AGACGAGTGTGCGGTGTGTGGAGCGCGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
Db	1346 AGGATTACTGTGAGAGTGTGCGCAGCAGGATGGGGAAATTAATCTGTGTGACACCTGCCCTC 1405
Oy	592 GGGCCTTCCACCTGGCCCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGGACCTGGA 651
Db	1406 GTGCTTACCACTGTGTGCTTGTGATCTTGAAGCTTACCGGAGCTTCAAGAGGCAATGGA 1465
Oy	652 GGTGCTTCAGCTGTGCTTCAGGCAACAGTCCAGAGAGTGCAGCCCCGGGACAGGA 706
Db	1466 GCTCCCTCACTGTGAGAGAGAGGGGGTCCAGTGTGGAGGCGCAAGAGAGAAAGAA 1520
RESULT 29	
AAD54629	
ID	AAD54629 standard; DNA; 6331 BP.
XX	
AC	AAD54629;
XX	
XX	26-JUN-2003 (first entry)
DE	Human chromodomain helicase DNA binding protein (CHD) encoding DNA #4.
XX	
KW	Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KM	chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KM	cell proliferative disorder; chromatin organisation modifier domain;
XX	cytosolic; ds.
XX	
OS	Homo sapiens.
XX	
PN	MO200298899-A2.
XX	
PD	12-DEC-2002.
XX	
PF	03-JUN-2002; 2002WO-US017466.
XX	
PR	05-JUN-2001; 2001US-0296076P.
PR	10-OCT-2001; 2001US-038605P.
PR	22-OCT-2001; 2001US-0338733P.
PR	15-FEB-2002; 2002US-0357253P.
PR	15-FEB-2002; 2002US-0357600P.
XX	
PA	(EXEL-) EXELIXIS INC.
XX	
PI	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI	Lioubin MM;
XX	
DR	WPI; 2003-156840/15.
XX	
PT	Identifying a candidate p53 pathway-modulating agent as therapeutic
PT	targets for disorders related to defective p53 function e.g. cancer by
PT	contacting an assay system having purified CHD polypeptide or nucleic
PT	acid, with a test agent.
XX	
PS	Disclosure; Page 51-54; 278pp; English.
CC	The present invention relates to a method for identifying candidate p53







XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-1B001105.  
XX 22-JUN-2001; 2001WO-1B001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
XX Katagiri F, Qian S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
XX pathogenic infection for conferring resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.  
XX  
XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
SQ  
Query Match 3.8%; Score 58.8; DB 8; Length 2000;  
Best Local Similarity 11.1%; Pred. No. 0.0067;  
Matches 80; Conservative 338; Mismatches 290; Indels 10; Gaps 4;  
QY 653 GTGCTCCAGTCCCTGCGAGCAACGATGAGAGTGAGAGCCCGGAGAGAGCCCGG 712  
DB 720 KTKYASASARWTRAKSRYSYRRRWYKRGWYRYRYRWSCHWTRAKSRKRWASAK 661  
QY 713 GCCCAGAGCAGCCCGTGGAGACCCCGG-TCCCCCGGCGCTAGTGGCGGAGAG 771  
DB 660 SCWMTWRGRSMWYKSCSAKCCKTRWYTSYSTGYMSYSKMSWTSKMSYWK 601  
QY 772 AGTAAAGAGTCCAGCTGGGAAACCCCTAGCCGAGTGAACAGACTTGTCTACAAGC 831  
DB 600 MTCMTWTSKSGSTRSKGRMSGMSRMWRRWKKRKYMKKCTWRRCMCRRMGY 541  
QY 832 ACCGCGCGCTCCGCTTGTGAGAGCCCGCTGCGAGTGTGAGCTTCCGCGCTGACCC 891  
DB 540 TWTTSRSRSMWYTKRYKRYTKRMYKRYKRYCYYYGYMKSYSMMRKYGCKACK 481  
QY 892 CCTCTAGTGTGGGTCTCTGAGAGGTGACAGAACTGCTGCTGGCGGCTGGGG 951  
DB 480 CCYAAACWKAAYSGMMWYTKYSKMRMSKTYMSMWM- - -YKCRSMYTGAKGYGCK 426  
QY 952 TGTGCGAGATGTATGAGAGTGTGCGGTGTACTGACCGCGCTGCTTCACTGAC 1011  
DB 425 MWTYSYGVKMYWYTWGSKYKSRCYKMR-MWYKMYMMWYSAVSSMMWYVYAKY 367  
QY 1012 GGTGCACTCCAGCGGAGCTCCCGCGGAGCGGAGCGCTGCGCTGACATCTGCT 1071  
DB 366 KTWYKRRGTSMWYKSKYKRYCTWCMYCMRCYRMRKMKRKTYSKRCYCRVATCW 307  
QY 1072 CAGAGAGTGAACCCAGCCCGCTGTGAGAGGGGTGTGAGCCCGCGCGCGCTGG 1131  
DB 306 CCCYRGRGYSRSRSMRTGAKMRMSWMSWRCYSYWKYKMYKMSYSYMSVMASSGT 247  
QY 1132 CCCCTGGGCTGCGAAGATGACACTGCGAGTGCAGAGCCCGCTGCAAGAGATGACC 1191  
DB 246 WBSRAKRYTKGYSTSRBRAMMRACRMYSACRYSRTSYCGSGYSGSKMYKMSKSCM 187

QY 1192 TGGAGTCCCTTTGAGCGAGCAGCACTTCGATGCGATCTGCGAGTGGCGCATCCAGAGCA 1251  
DB 186 - - -RMTCSWCSCTCYCYAGACMSWMSGCGCTRWKMRSKYSKMSCKKXGSCC 130  
QY 1252 TGGCGCGTCCGCGGCGCCCTTCCCTCTGAGCCCGAGATGCGCGGAGCATGAGCTGCG 1311  
DB 129 TKYCSITGYRYCKMYKYSYKCYCTWMSYKRMKMSKMSKMSKMSKMSKMSKMSKMSK 70  
QY 1312 ATGAGAGATGCTGGAAGACACCTCTCTCTGAGTCTGAGAGCCCGCGCTGCG 1369  
DB 69 TSRMWSMYAAMKMGSGSMYRMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSK 12  
RESULT 34  
AAK53491  
ID AAK53491 standard; DNA; 114955 BP.  
XX  
XX AAK53491;  
AC  
XX  
XX 05-JUL-1999 (first entry)  
DT  
XX  
XX Human adenosine A1 receptor antisense oligonucleotide fragment.  
DE  
XX  
XX Antisense oligonucleotide; multiple target; antisense treatment;  
KW impaired respiration; inflammation; lung disease;  
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
KW acute asthma; allergy; asthma; impaired respiration;  
KW respiratory distress syndrome; pain; cystic fibrosis;  
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; se.  
KW  
XX  
XX Synthetic.  
OS  
XX  
XX WO9313886-A1.  
PN  
XX  
XX 25-MAR-1999.  
PD  
XX  
XX 17-SEP-1998; 98WO-US019419.  
PF  
XX  
XX 17-SEP-1997; 97US-0059160P.  
PR  
XX  
XX 09-JUN-1998; 98US-00093972.  
PR  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA  
XX  
XX Nyce JW;  
PI  
XX  
XX WPI; 1999-229400/19.  
DR  
XX  
XX New antisense oligonucleotides used in treatment of, e.g. pulmonary  
PT vasoconstriction.  
PT  
XX  
XX Disclosure; Page 37; 120pp; English.  
PS  
XX  
XX The specification describes antisense oligonucleotides (AAK52869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene initiation  
CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-  
CC end and the juxta-section between coding and non-coding regions and all  
CC segments of RNAs encoding proteins associated with one or more diseases,  
CC conditions or mixtures. The antisense oligonucleotides may be derived  
CC from sequences AAK55180-271. These multiple target oligonucleotides  
CC (specifically AAK55180-271) can be used for the antisense treatment of  
CC diseases and conditions. Typical diseases and conditions are those  
CC associated with impaired respiration and inflammation, including lung  
CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,  
CC acute asthma, allergies, asthma, impaired respiration, respiratory  
CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,  
CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary  
CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.

CC colon cancer, breast cancer, lung cancer, pancreatic cancer,  
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as  
CC well as all types of cancers which may metastasize or have metastasized  
CC to the lungs, including breast and prostate cancer  
XX  
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 3.7%; Score 57.4; DB 2; Length 114955;

Best Local Similarity 29.7%; Pred. No. 0.023; Indels 5; Gaps 2;

Matches 301; Conservative 104; Mismatches 604; Indels 5; Gaps 2;

QY 257 CGGGGCCCCCTGGAGACGACGACCTGCAAGAAACCGGGTTTCTTCCCAATAGGATGAC 316  
DB 104081 CCGGGCCSNNNDNNCCGCGTGGGGCCGGGCSNNNDNNGGCTGGGGCCGGGCSNNNDNNCC 104140  
QY 317 CCGGGGGGTGTCTGTTCGAGACAGAGTGAATGGGACAGGTGCTCAGGGCAAAATTC 376  
DB 104141 GTCGGGCCGGGCSNNNDNNCGTCGGGCGGGCSNNNDNNGTGGGCGGGCSNNNDNNTC 104200  
QY 377 AGGCTCTGGACATGGAGCAGAGGCGAGAGACTGGGGAACTTCAAGTACCAGATGCTG 436  
DB 104201 GGGCCGGGCSNNNDNNCGGCGCSNNNDNNCCGCBGCGCGG---GCCGCGCGCG 104256  
QY 437 CTGGGGAGAGCTGTTTGGAGAGAGGTGCTCTCAGAGAGGTGCTGCAACCCAGCCAGT 496  
DB 104257 GCGGGGCSNNNDNNCCGCBGCGGGCGGCGCGCGCGGCGGCSNNNDNNCCGCBG 104316  
QY 497 CTGCATGGGCGTCTCTTCTGCTGCTGCAAGAAATGAGACGAGTGTGCTGTGTGGGA 556  
DB 104317 GCBGGGGCGCGCGCGCGGCGGCSNNNDNNCCGCBGCGCGCGCGCGCGCGCGG 104376  
QY 557 GCGCGGGAGCTCATCTGCTGTGAGCGCTGCGCGCTTGCACCTGGCGTGCCTGTC 616  
DB 104377 SNNNDNNCCGCGCGCGCGCGCGCGCGCGCGGCSNNNDNNCCGCBGCGCGCGC 104436  
QY 617 CCCTCCGCTCCGCGAGATCCCAAGTGGAGACTGAGGTGCTCAGTGCCTGCAAGCAAC 676  
DB 104437 CCGCGGCGSNNNDNNCCGCBGCGCGGCGCGCGCGCGGCSNNNDNNCCGCBGCGCG 104496  
QY 677 AGTCCAGAGGTGACACCCCGGCAAGAGACCCCGCCCAAGAGACCCGTGAGAC 736  
DB 104497 GCGCGCGCGCGCGGCSNNNDNNCCGCBGCGCGCGCGCGGCSNNNDNNCCGCBGCGC 104556  
QY 737 CCGGCTCCCGCGGGCTTAAGTGGCGGGAGAGAGTAAAGTCCACTGGGGAGCC 796  
DB 104557 CCGGGCGCGCGCGGCSNNNDNNCCGCBGCGCGCGCGCGGCSNNNDNNCCGCBGCGC 104616  
QY 797 CCTAGCGGCAATGACACGACTCTTGTCTTACAGACACTGCGCGCTCCGCTTCTGACAC 856  
DB 104617 BGGGGCGCGCGGCSNNNDNNCCGCBGCGCGGCGCGCGGCSNNNDNNCCGCBGCGCGC 104676  
QY 857 CCGGCTGCGAGTGTGAGATCTCTCGGCGCTGCAACCCCTTACTGTGTGGGTCTTGAAGG 916  
DB 104677 GCGGCSNNNDNNCCGCBGCGCGGGCGCGGCSNNNDNNCCGCBGCGCGGCSNNNDN 104736  
QY 917 TCAGCAGACACTGGCTCCTGTGTGCGGCTGCGGGGTGTCGAGATGATAGCACTGCT 976  
DB 104737 NCCGCBGCGCGCGGCSNNNDNNCCGCBGCGCGGCGGCSNNNDNNCCGCBGCGCGGCG 104796  
QY 977 GCGGTGTACTCACTGCGCGCTTCTTCACTGGCGCTGCCACTTCCAGCGCGCACTTC 1036  
DB 104797 SNNNDNNCCGCBGCGCGGCSNNNDNNCCGCBGCGCGGCSNNNDNNCCGCBGCGCGSNN 104856  
QY 1037 CCGGCGCGGAGAGGCGCTGCGTGTGCAATCTGTCTAGAGAGAGTAAACCCAGCCCTGT 1096  
DB 104857 NDNCCGCBGCGCBSNNNDNNCCGCBGCGCGGCSNNNDNNCCGCBGCGCGGCGCGCGCG 104915  
QY 1097 GGAGGGGGGTGAGCGCGCGCGCGCGCGCTGCGCGCTGCGCAAGATGACAC 1156  
DB 104916 GCGGCGGCSNNNDNNCCGCBGCGCGCGCGCGCGCGCGGCGGCSNNNDNNCCGCBGCGCB 104975  
QY 1157 TCCCAATCAAGACCGCGCTCTGCAAGGAGATGACTGAGTCCCTTCTGAGCGAGACAC 1216

DB 104976 GGGCGCGCGCGCGCGGCGGCSNNNDNNBGGCCBGGCGCGCGCGCGGCGGCSNN 105035  
QY 1217 CTTCATGAGCATCTCTGACGTGGGCGCATTCAGAGATGCGCGCTCGGGGCGGCCCC 1270  
DB 105036 NDNNGCCBGGGCGCGCGCGCGCGGCGGCSNNNDNNCCBGGGCGCGCGCGC 105089

RESULT 35

AB109061  
ID AB109061 standard; cDNA; 6124 BP.

AC AB109061;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21665.  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
XX  
XX PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX  
XX MPI; 2001-656860/75.  
XX  
XX P-PSDB; ABB64958.  
XX  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
XX PS Claim 1; SEQ ID NO 21665; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signaling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101640-AB16175) and the encoded proteins (AB57737-  
XX AB872072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 6124 BP; 1706 A; 1562 C; 1509 G; 1347 T; 0 U; 0 Other;

Query Match 3.7%; Score 56.4; DB 4; Length 6124;  
Best Local Similarity 59.3%; Pred. No. 0.025; Indels 0; Gaps 0;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 528 AATGAGACGAGTGTGCTGTGCGGACGCGGAGCTCATCTGTGACGAGCTGC 587  
DB 3391 AACGAGAGTGTGCGCGCTGTGTGATGAGCGAGCTATGTGCTGCGACAAATGT 3450  
QY 588 CCTGGGCGCTTCCACCTGCGCTGCTGCTCCCTCGCTCCGGAGATCCCACTGGAGCC 647  
DB 3451 CCCAAAGTTTTCATCAAGAACTGTCAATCCCTGAGATCACTGCTTGGCGGACGAGAC 3510  
QY 648 TGAAGTGTCTCAGTGCCTGAGGACACAGTCCAGAGGTG 689  
DB 3511 GAGAGCTGCAAGTGTCTGCTGCTTCAATCAAGAGAGCTG 3552





CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (II) is useful in gene therapy techniques to restore normal  
CC activity of (I) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

CC SQ Sequence 2495 BP; 696 A; 581 C; 733 G; 485 T; 0 U; 0 Other;  
CC  
CC Query Match 3.6%; Score 55.4; DB 5; Length 2495;  
CC Best Local Similarity 62.8%; Pred. No. 0.037;  
CC Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGGCTGCCCTC 591  
DB 1367 AGGACATATTCGACAGGTGTGCGGACGAGCGGTGATCATCTGTGTACTGTGCCCC 1426  
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTTGA 651  
DB 1427 GTGCTTACCAACATGTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCAAGTGA 1486  
QY 652 GGTGCTCCAGCTGCCCTG 668  
DB 1487 GCTGCCACACCTGCGTG 1503

RESULT 40  
AA574874  
ID AA574874 standard; cDNA; 7132 BP.  
AC AA574874;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #10678.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Dmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG10687.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostic, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID NO 10678; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (II) is useful in gene therapy techniques to restore normal  
CC activity of (I) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

CC SQ Sequence 7132 BP; 1723 A; 1875 C; 1964 G; 1569 T; 0 U; 1 Other;  
CC  
CC Query Match 3.6%; Score 55.4; DB 5; Length 7132;  
CC Best Local Similarity 62.8%; Pred. No. 0.042;  
CC Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGGCTGCCCTC 591  
DB 1367 AGGACATATTCGACAGGTGTGCGGACGAGCGGTGATCATCTGTGTACTGTGCCCC 1426  
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTTGA 651  
DB 1427 GTGCTTACCAACATGTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCAAGTGA 1486  
QY 652 GGTGCTCCAGCTGCCCTG 668  
DB 1487 GCTGCCACACCTGCGTG 1503

RESULT 41  
AAK71358  
ID AAK71358 standard; DNA; 37314 BP.  
AC AAK71358;  
XX  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26170.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX

PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251038P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM,  
PI  
XX  
DR WPI; 2001-483426/52.  
XX  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
PS Disclosure; SEQ ID NO 26170; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC



CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 37314 BP; 8700 A; 8411 C; 9616 G; 10587 T; 0 U; 0 Other;

Query Match 3.6%; Score 55.4; DB 4; Length 37314;  
Best Local Similarity 62.8%; Pred. No. 0.053;  
Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGCTCATCTGCTGTGACGGCTGCGCTC 591  
DB 7046 AGGACTATTGCGAGGTGTGCGACGAGCGGTGATCATCTGCTGTGATCTGTCTCC 7105  
QY 592 GGGCTTCCACCTGGCTGCTGCTGCTCCCTCCCTCCGGAGATCCCAAGTGGACCTGGA 651  
DB 7106 GTGCTTACCACTGTGTGCTGCTGTGATCCGACATGAGAGGCTCCCGAGGCAAGTGA 7165  
QY 652 GGTGCTCCACGCTGCTG 668  
DB 7166 GCTGCCACACCTGCGTG 7182

RESULT 42  
ABO88207/c  
ID ABO88207 standard; cDNA; 172570 BP.

XX ABO88207;

XX 18-SEP-2002 (first entry)

XX Human osteoblast differentiation related cDNA SEQ ID NO 114.

XX Human; osteoblast; stem cell differentiation; bone tissue deposition;  
XX osteoporosis; osteopathic; ss.

XX Homo sapiens.

XX WO200250301-A2.

XX 27-JUN-2002.

XX 18-DEC-2001; 2001WO-US048276.

XX 18-DEC-2000; 2000US-0255882P.

XX 24-APR-2001; 2001US-0285691P.

XX (GENE-) GENE LOGIC INC.  
XX (PROC ) PROCTER & GAMBLE CO.

XX Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
XX Mertz U;

XX MPI; 2002-557663/59.

XX Use of genes and their expression profiles associated with osteoblast  
XX differentiation for screening modulators bone formation, for diagnosing  
XX or treating e.g. osteoporosis, or as markers for the differentiation  
XX process.

XX Claim 1, SEQ ID NO 114; 78bp + Sequence Listing; English.

CC The invention relates to genes and their expression profiles are used  
CC for: (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
CC deposition of bone tissue, abnormal rate of osteoblast formation or  
CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
CC cited in (b), or monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
CC induced abnormalities in bone formation or bone loss, conditions that  
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
CC or fibrous dysplasia. The present sequence is that of an osteoblast  
CC differentiation associated cDNA marker of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences

Query Match 3.6%; Score 55.4; DB 6; Length 172570;  
Best Local Similarity 62.8%; Pred. No. 0.066;  
Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGCTCATCTGCTGTGACGGCTGCGCTC 591  
DB 157430 AGGACTATTGCGAGGTGTGCGACGAGCGGTGATCATCTGCTGTGATCTGTCTCC 157371  
QY 592 GGGCTTCCACCTGGCTGCTGCTGCTCCCTCCCTCCGGAGATCCCAAGTGGACCTGGA 651  
DB 157370 GTGCTTACCACTGTGTGCTGCTGTGATCCGACATGAGAGGCTCCCGAGGCAAGTGA 157311  
QY 652 GGTGCTCCACGCTGCTG 668  
DB 157310 GCTGCCACACCTGCGTG 157294

RESULT 43

XX AAA45365  
XX ID AAA45365 standard; cDNA; 523 BP.

XX AAA45365;

XX 21-AUG-2000 (first entry)

XX Mouse secreted expressed sequence tag SEQ ID NO:1940.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;  
XX expressed sequence tag; EST; probe; chemotactic; proliferative;  
XX immunomodulatory; haematopoietic; chemokine; angiogenic; haemostatic;  
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
XX antiviral; antidiabetic; antiasclerotic; vulnery; antiparkinsonian;  
XX antitumor; osteopathic; neuroprotective; nootropic; antiparkinson;  
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;  
XX autoimmune disorder; multiple sclerosis; allergic condition;  
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
XX central nervous system disorder; Alzheimer's disease; stroke;  
XX Parkinson's disease; Huntington's disease; coagulation disorder;  
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;  
XX infection; depression; psoriasis; ss.

XX Mus musculus.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US024206.

XX 15-OCT-1998; 98US-0104436P.

XX (GENE ) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Metberg D, Treacy M, Bowman MR;  
XX WPI: 2000-317938/27.  
XX  
XX Isolated polynucleotides, and encoded proteins, comprising secreted  
PT expressed sequence tags (ESTs), useful for treating various disorders  
PT such as autoimmune, infectious, and central nervous system disorders.  
XX  
XX Claim 1, Page 695; 803pp; English.  
XX  
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed  
CC sequence tags (ESTs), isolated from human, mouse, chicken and rat tissue  
CC sources. The ESTs can have a range of activities depending on the  
CC tissues they were isolated from. The activities include: chemotactic;  
CC proliferative; immunomodulatory; haematopoietic; chemokine; analgesic;  
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;  
CC antifungal; antiviral; antidiabetic; antilasthmatic; vulnery; anticancer;  
CC osteopathic; neuroprotective; nootropic; antiparkinsonian; antiporiotic;  
CC cerebroprotective; anticonvulsant; and antidepressant. The ESTs can be  
CC used for gene therapy and in vaccines. The ESTs are useful as probes for  
CC the identification and isolation of full-length cDNAs and genomic DNA  
CC molecules which correspond to the ESTs. Proteins encoded by the ESTs  
CC are useful in assays for determining biological activity and raising  
CC antibodies. They may be useful for treatment of autoimmune disorders  
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
CC osteoporosis, osteoarthritis, central nervous system disorders  
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
CC disease), tumours, bacterial, fungal or viral infections, depression and  
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
CC in the exemplification of the present invention  
XX  
XX Sequence 523 BP; 134 A; 113 C; 167 G; 96 T; 0 U; 13 Other;  
SQ  
Query Match 3.6%; Score 55.2; DB 3; Length 523;  
Best Local Similarity 59.6%; Pred. No. 0.033;  
Matches 93; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 532 AGGACGAGTGTGCTGCTGGGACGCGGGAGCTCATCTGCTGACGCGCTCC 591  
DB 186 AGGACTATTGCGAGTGTGCGACGAGGAGGAGATCATCTGCTGATCCTGTC 245  
QY 592 GGGCTTCACCGCTGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTGA 651  
DB 246 GAGCTACCATATGATGCTGCTGAGACCAAGATGAGAGAGGCCCGAGGGCAAGTGA 305  
QY 652 GGTGTCTCAGCTGCTGCGAGGCAACAGTCCAGAGG 687  
DB 306 GCTGTCCCACTGTGAGAGGAGGAGATCCAGTGG 341  
RESULT 44  
ACH48320  
ID ACH48320 standard; cDNA; 494 BP.  
XX  
XX ACH48320;  
XX  
XX 13-OCT-2003 (first entry)  
XX  
XX Human lung tumour cDNA #453.  
XX  
XX DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
XX genome mapping; biodiversity; genetic disorder.  
XX  
XX OS Homo sapiens.  
XX  
XX PN US2003073623-A1.  
XX  
XX PD 17-APR-2003.  
XX  
XX PF 30-JUL-2001; 2001US-00918995.

XX  
XX 30-JUL-2001; 2001US-00918995.  
XX  
XX (DRMA/) DRMANAC R T.  
XX  
XX (LABA/) LABAT I.  
XX  
XX (STAC/) STACHE-CRAIN B.  
XX  
XX (DICK/) DICKSON M C.  
XX  
XX (JONE/) JONES L W.  
XX  
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX  
XX WPI: 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
XX Claim 1, SEQ ID NO 35532; 44pp; English.  
XX  
XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnosis as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC segdata.uspto.gov/sequence.html?docID=20030073623  
XX  
XX Sequence 494 BP; 154 A; 101 C; 102 G; 136 T; 0 U; 1 Other;  
SQ  
Query Match 3.5%; Score 54.2; DB 9; Length 494;  
Best Local Similarity 61.9%; Pred. No. 0.053;  
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 528 AATGAGGACGAGTGTGCTGCTGCGGAGCGGAGCTCATCTGCTGAGCGGCTGC 587  
DB 109 AATGAGGAGTGTGTGCTGCTGCAAAAGGAGGAACTCTCTGCTGAAAAATGTC 168  
QY 588 CCTCGGACCTTCACCTGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCC 647  
DB 169 CCCAAGATATTCATCTTCTGTGATGAGCCACATTTGACAAATTTTCCAGTGGAGAG 228  
QY 648 TGGAGTGTCTCAAGTCC 666  
DB 229 TGGATTTGCACTTCTGCC 247  
RESULT 45  
ADB80995  
ID ADB80995 standard; DNA; 3039 BP.  
XX  
XX ADB80995;  
XX  
XX 04-DEC-2003 (first entry)  
XX  
XX RING-SH complex related DNA, SEQ ID No 69.  
XX  
XX DE RING-SH 3; Gag protein; Gag late domain; PI3K; actin; myosin; Hsp60;  
XX Hsp70; Hsp90; STM1; STAM2a; STAM2b; VHS-UIM; GTPase; E2 enzyme; ts9101;  
XX cullin; RING-SH; clathrin; v-src; vaccine; antiviral; retrovirus;  
XX rhadovirus; filovirus; gene; ds.  
XX  
XX OS unidentified.

XX WO2003033646-A2.  
PN  
XX  
XX 24-APR-2003.  
PD  
XX 31-JUL-2002; 2002WO-US024589.  
PF  
XX 31-JUL-2001; 2001US-0308958P.  
XX PR 09-NOV-2001; 2001US-0345846P.  
XX  
XX (PROT-) PROTEOLOGICS INC.  
PI Greener T, Moskowitz H, Reiss Y, Alroy I;  
PI WPI; 2003-393509/37.  
DR P-PSDB; ADB80954.  
XX  
XX New isolated protein complex comprising a RING-SH 3 polypeptide and  
PT another polypeptide, useful for detecting cells infected with a virus,  
PT and for treating viral disorders caused by retroviruses, rhadoviruses,  
PT or filoviruses.  
XX  
XX Disclosure; Fig 68; 176pp; English.  
PS  
XX The invention relates to a novel isolated protein complex comprising a  
CC RING-SH 3 polypeptide and a polypeptide selected from a Gag protein, a  
CC Gag late domain, PI3K, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2,  
CC STAM2B, VHS-UM, a GTPase, an E2 enzyme, teg101, a cullin, RING-SH, and a  
CC clathrin. The novel protein complex has virucidal activity and can be used  
CC to treat disorders as part of a vaccine. The protein complex and  
CC composition are useful for detecting cells infected with a virus, for  
CC identifying agents having antiviral activity, and for treating viral  
CC disorders caused by retroviruses, rhadoviruses, or filoviruses. This  
CC polynucleotide represents a DNA sequence relating to a protein comprising  
CC the RING-SH complex of the invention.  
XX  
XX Sequence 3039 BP; 941 A; 738 C; 664 G; 696 T; 0 U; 0 Other;  
SQ  
Query Match 3.5%; Score 54.2; DB 9; Length 3039;  
Best Local Similarity 61.9%; Pred. No. 0.068;  
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 528 AATGAGGAGAGTGTGCGGTGTCGGACGGGGGAGCTATGCTGACGGCTGC 587  
DB 2359 AATGAGGAGTGTGTCAGTTGTCAAAACGAGGGGAACTCTCTGCTGTGAAAAGTGC 2418  
QY 588 CCTCGGGCCCTTCACCTGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAATGGGACC 647  
DB 2419 CCCAAGTATTCATCTTCTTCTCATGTGCCCATTTGACAAATTTTCCAAAGTGGAGAG 2478  
QY 648 TGGAGGTGCTCCAGCTGCC 666  
DB 2479 TGGATTTCACCTTTCTGCC 2497  
RESULT 46  
ADJ56478  
ID ADJ56478 standard; cDNA; 3917 BP.  
XX  
XX ADJ56478;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX Human cDNA differentially expressed in MYCN activated cells SeqID 284.  
DE  
XX  
XX human; differential expression; transactivator; proto-oncogene;  
KW neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;  
XX MYCN activated cell.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX US2003119009-A1.  
XX

PD 26-JUN-2003.  
XX  
XX 25-FEB-2002; 2002US-00084817.  
XX PF  
XX 23-FEB-2001; 2001US-0270784P.  
XX PR  
XX (STUN/) STUART S G.  
XX PA (NUCH/) NUCHTERN J G.  
XX PA (PLON/) PLON S E.  
XX PA (SHOH/) SHOHET J M.  
XX  
XX Stuart SG, Nuchtern JG, Plon SE, Shohet JM;  
XX WPI; 2003-635698/60.  
XX  
XX New genes regulated by MYCN activation, useful in gene therapy,  
PT particularly for treating a subject with e.g. neuroblastoma or other  
PT cancers, or for diagnosing, staging or monitoring the treatment of the  
PT cancer.  
XX  
XX Claim 1; SEQ ID NO 284; 27pp; English.  
PS  
XX This invention relates to novel isolated cDNAs that are differentially  
CC expressed in MYCN activated cells. Specifically, it refers to  
CC polynucleotide sequences that exhibit differential expression patterns in  
CC cells activated by the transactivator MYCN, where MYCN is a proto-  
CC oncogene that is amplified in neuroblastoma cells and is common in small  
CC cell lung cancers. The present invention describes these cDNA molecules  
CC as useful for in hybridisation assays to detect expression of nucleic  
CC acids (or complementary nucleic acids) in a present in a given sample, as  
CC well as for screening assays by identifying molecules or compounds that  
CC specifically bind the cDNA as a ligand and modulate function or activity.  
CC Accordingly, these compositions exhibit cytostatic activity and can also  
CC be used for gene therapy purposes. This polynucleotide sequence is a cDNA  
CC that is differentially expressed in MYCN activated cells, given in an  
CC exemplification of the invention. NOTE: This sequence does not appear in  
CC the printed specification but has been obtained in electronic format from  
CC the US Patent Office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID=20030119009.  
XX  
XX Sequence 3917 BP; 1232 A; 863 C; 843 G; 979 T; 0 U; 0 Other;  
SQ  
Query Match 3.5%; Score 54.2; DB 10; Length 3917;  
Best Local Similarity 61.9%; Pred. No. 0.071;  
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 528 AATGAGGAGAGTGTGCGGTGTCGGACGGGGGAGCTATGCTGTCGACGGCTGC 587  
DB 2359 AATGAGGAGTGTGTCAGTTGTCAAAACGAGGGGAACTCTCTGCTGTGAAAAGTGC 2418  
QY 588 CCTCGGGCCCTTCACCTGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAATGGGACC 647  
DB 2419 CCCAAGTATTCATCTTCTTCTCATGTGCCCATTTGACAAATTTTCCAAAGTGGAGAG 2478  
QY 648 TGGAGGTGCTCCAGCTGCC 666  
DB 2479 TGGATTTCACCTTTCTGCC 2497  
RESULT 47  
AAK94864  
ID AAK94864 standard; cDNA; 3997 BP.  
XX  
XX AAK94864;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human full-length cDNA, SEQ ID NO: 4044.  
DE  
XX  
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX

PN EP130094-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114089.  
 XX  
 PR 08-JUL-1999; 99JP-00194486.  
 XX 11-JAN-2000; 2000JP-00118774.  
 PR 02-MAY-2000; 2000JP-00183765.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX  
 DR WPI; 2001-524255/58.  
 DR F-PSDB; AAM93902.  
 XX  
 PT 830 Primers useful for synthesizing full length cDNA clones and their use  
 PT in genetic manipulation.  
 XX  
 PS Claim 8; SEQ ID NO 4044; 1380bp + Sequence Listing; English.  
 XX  
 CC The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been isolated  
 CC and nucleotide sequences of 5' - and 3' - ends of the cDNA molecules have  
 CC been determined. Primers for synthesizing the full length cDNA are useful  
 CC for clarifying the function of the protein encoded by the cDNA. The full  
 CC length clones were obtained by construction of full length enriched cDNA  
 CC libraries that were synthesized by the oligo-capping method. The primers  
 CC enable the production of the full length cDNA easily without any special  
 CC methods. The present sequence is a full length human cDNA of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in CD-ROM format directly  
 CC from EPO  
 CC  
 SQ Sequence 3997 BP; 1221 A; 944 C; 911 G; 921 T; 0 U; 0 Other;  
 XX  
 Query Match 3.5%; Score 54.2; DB 4; Length 3997;  
 Best Local Similarity 61.9%; Pred. No. 0.071;  
 Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
 QY 528 AATGAGACGAGTGTGCGGTGTCGAGACGGCGGAGCTCATCTGTGTGACGGCTGC 587  
 DB 2688 AATGAGACGAGTGTGCGGTGTCGAGTGTCAAAACGAGGGGAATCTCTGCTGTGAAAAGTGC 2747  
 QY 588 CCTCGGCGCTTCCACTGAGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAAGTGGAC 647  
 DB 2748 CCCAAAGTATTCATCTTCTGTGATGTGCCCCACATTGACAAATTTCCAAGTGGAG 2807  
 QY 648 TGGAGGTGCTCCAGCTGCC 666  
 DB 2808 TGGATTGCACTTTCTGCC 2826  
 XX  
 RESULT 48  
 ADL32011  
 ID ADL32011 standard; cDNA; 3997 BP.  
 XX  
 AC ADL32011;  
 XX  
 DT 20-MAY-2004 (first entry)  
 XX  
 DE Full length human cDNA clone SegID 4044.  
 XX  
 KW human; medicine; signal transduction; glycoprotein; transcription;  
 KW oligo-capping method; ss; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1396543-A2.  
 XX  
 PD 10-MAR-2004.  
 XX

XX  
 PF 07-JUL-2000; 2003EP-00025638.  
 XX  
 PR 08-JUL-1999; 99JP-00194486.  
 XX 11-JAN-2000; 2000JP-00118774.  
 PR 02-MAY-2000; 2000JP-00183865.  
 PR 07-JUL-2000; 2000EP-00114089.  
 XX  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX  
 DR WPI; 2004-204755/20.  
 DR F-PSDB; ADL32012.  
 XX  
 PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full  
 PT length human cDNAs.  
 XX  
 PS Example 1; SEQ ID NO 4044; 1340bp; English.  
 XX  
 CC This invention relates to a novel primers useful for synthesizing full  
 CC length cDNA molecules that encode human proteins. Specifically, it refers  
 CC to secretory or membrane proteins that are potential therapeutic agents/  
 CC target molecules in the field of medicine, and in particular genes  
 CC encoding proteins that are associated with signal transduction,  
 CC glycoproteins and transcription. The present invention describes a method  
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'  
 CC ends using the oligo-capping method. This polynucleotide sequence is a  
 CC full length human cDNA clone of the invention.  
 CC  
 SQ Sequence 3997 BP; 1221 A; 944 C; 911 G; 921 T; 0 U; 0 Other;  
 XX  
 Query Match 3.5%; Score 54.2; DB 12; Length 3997;  
 Best Local Similarity 61.9%; Pred. No. 0.071;  
 Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
 QY 528 AATGAGACGAGTGTGCGGTGTCGAGACGGCGGAGCTCATCTGTGTGACGGCTGC 587  
 DB 2688 AATGAGACGAGTGTGCGGTGTCGAGTGTCAAAACGAGGGGAATCTCTGCTGTGAAAAGTGC 2747  
 QY 588 CCTCGGCGCTTCCACTGAGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAAGTGGAC 647  
 DB 2748 CCCAAAGTATTCATCTTCTGTGATGTGCCCCACATTGACAAATTTCCAAGTGGAG 2807  
 QY 648 TGGAGGTGCTCCAGCTGCC 666  
 DB 2808 TGGATTGCACTTTCTGCC 2826  
 XX  
 RESULT 49  
 AAX53491/C  
 ID AAX53491 standard; DNA; 114955 BP.  
 XX  
 AC AAX53491;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
 XX  
 KW Antisense oligonucleotide; multiple target; antisense treatment;  
 KW impaired respiration; inflammation; lung disease;  
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 XX  
 OS Synthetic.  
 XX

PN MO9913886-A1.  
 XX 25-MAR-1999.  
 XX 17-SEP-1998; 98WO-US019419.  
 XX 17-SEP-1997; 97US-0059160P.  
 PR 09-JUN-1998; 98US-00093972.  
 XX (UYEC-) UNIV EAST CAROLINA.  
 XX PA  
 XX NYce JW;  
 XX WPI; 1999-229400/19.  
 DR  
 XX  
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction.  
 XX  
 PS Disclosure; Page 37; 120pp; English.  
 CC The specification describes antisense oligonucleotides (AA52869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene initiation  
 CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-  
 CC end and the juxta-section between coding and non-coding regions and all  
 CC segments of RNAs encoding proteins associated with one or more diseases,  
 CC conditions or mixtures. The antisense oligonucleotides may be derived  
 CC from sequences AA55272-74. These multiple target oligonucleotides  
 CC (specifically AA55180-271) can be used for the antisense treatment of  
 CC diseases and conditions. Typical diseases and conditions are those  
 CC associated with impaired respiration and inflammation, including lung  
 CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,  
 CC acute asthma, allergies, asthma, impeded respiration, respiratory  
 CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,  
 CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary  
 CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.  
 CC colon cancer, breast cancer, lung cancer, pancreatic cancer,  
 CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as  
 CC well as all types of cancers which may metastasize or have metastasized  
 CC to the lungs, including breast and prostate cancer  
 XX  
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;  
 Query Match 3.5%; Score 54; DB 2; Length 114955;  
 Best Local Similarity 30.8%; Pred. No. 0.12;  
 Matches 226; Conservative 63; Mismatches 443; Indels 2; Gaps 1;

QY 548 GGTGGGGAGCGGGGAGATCATCTGTCAGCGCTGCGCTCCGACCTGAGC 607  
 DB 105539 GGGCGGAGCGAGGGGCGCCNNHNNNSVCGAGGCTVCGCGCCNNHNNNSCGGCGGCGCN 105480  
 QY 608 CTGCGCTGTCCTCCCTCCGCTCCGAGATCCCACTGAGGAGTGTCTCAGTCCCT 667  
 DB 105479 HNNNSCGGCGCGCGCGNNHNNNSCGGCGCGCGGCGGNNHNNNSCGGCGCGGCGGNNHN 105420  
 QY 668 GCAGGCAAGTTCAGAGAGTGCAGCCCGGCGGAGAGAGCCCGGCGCCAGAGGACAC 727  
 DB 105419 NSCGGCGCGCGCGCGNNHNNNSCGGCGCGCGGCGGNNHNNNSCGGCGCGCGCGG 105360  
 QY 728 CGTGGAGACCGCGCTCCCGCGGGGCTTGGTGGGCGGAGAGAGTAAAGAGTCCAC 787  
 DB 105359 CNNHNNNSCGGCGCGCGCGCGGCGGCGGCGGCGCGCGCGGCGGCGGCGGCGGCGG 105300  
 QY 788 TGGGGAACCCCTAGCGGAGATGACACGACTTGTCTACAAGACACTCTCGGCTCGGC 847  
 DB 105299 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105240  
 QY 848 TTCTGAGCGCGCGCTCCAGAGTGTGACTTC--TCGGCCCTGAGACCCCTACTGTGTG 905  
 DB 105239 GCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 105180  
 QY 906 GGTCTGAGAGGTGACAGAAACCTGCTCTGTGGTGGCGGTTGGCGGAGATGCT 965

DB 105179 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105120  
 QY 966 ACGAGCTGCTCGGCTGTAATCATGCGCGCTGCTTCACTGCGCTGCCACTTCCGA 1025  
 DB 105119 CCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105060  
 QY 1026 GCGGCGCTCCCGCGCGGAGCGGAGCGGCTGCTGCAATCTGTCTCAGAGACGAGAC 1085  
 DB 105059 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105000  
 QY 1086 CCAGCCCTGTGAGAGGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1145  
 DB 104999 NSCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104940  
 QY 1146 AAGATGACATGTCAGATCAGAGCGGCTGTGCAAGAGATGATCTGAGTCTCTTG 1205  
 DB 104939 VGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104880  
 QY 1206 AGCGAGACACTTGTGATGATGATCTCTGAGTGGGCGGCGGCGGCGGCGGCGGCGG 1265  
 DB 104879 GGNHNNNSVGGCGGCGGNNHNNNSVGGCGGCGGNNHNNNSVGGCGGCGGNNHN 104820  
 QY 1266 GCCCGCTTCCCTC 1279  
 DB 104819 NNSCCVGGCGGCGG 104806

RESULT 50  
 AA125090/c  
 ID AA125090 standard; DNA; 292 BP.  
 XX  
 AC AA125090;  
 XX  
 DT 12-OCT-2001 (first entry)  
 DE Probe #15023 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; 88.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200157278-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000670.  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human cervical epithelial cells.  
 XX  
 PS Claim 25; SEQ ID NO 15023; 487bp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging of

CC diseases of the cervix, notably cervical cancer. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;

Best Local Similarity 61.8%; Pred. No. 0.098; Mismatches 52; Indels 0; Gaps 0;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGATGGGGAATGCTCTGTGATACCTGTC 81

QY 589 CTCGGGCTTCCACTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGGACCT 648

DB 80 CTTCTTCTTACCACTCCACTGCTGAATCCCCCACTTCCAGAGATCCCAACGGTGAAT 21

QY 649 GGAGGTGCTCCAGCTG 664

DB 20 GGCTCTGTCCCGTTG 5

RESULT 51

ABA70776/c

ID ABA70776 standard; DNA; 292 BP.

AC ABA70776;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #19081.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS Claim 4; SEQ ID NO 19081; 639pp + Sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring

CC human gene expression in a sample derived from human foetal liver. The

CC single exon nucleic acid probes may be used for predicting, measuring and

CC displaying gene expression in samples derived from human foetal liver. The

CC present sequence is a single exon nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;

Best Local Similarity 61.8%; Pred. No. 0.098; Mismatches 52; Indels 0; Gaps 0;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGATGGGGAATGCTCTGTGATACCTGTC 81

QY 589 CTCGGGCTTCCACTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGGACCT 648

DB 80 CTTCTTCTTACCACTCCACTGCTGAATCCCCCACTTCCAGAGATCCCAACGGTGAAT 21

QY 649 GGAGGTGCTCCAGCTG 664

DB 20 GGCTCTGTCCCGTTG 5

RESULT 52

AA150949/c

ID AA150949 standard; DNA; 292 BP.

AC AA150949;

DT 17-OCT-2001 (first entry)

DE Probe #19635 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

OS genetic disorder; ss.

PN Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PS Claim 25; SEQ ID NO 19635; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENPs).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders

CC for antenatal diagnosis of human genetic disorders

CC for antenatal diagnosis of human genetic disorders

XX Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

XX Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;

Best Local Similarity 61.8%; Pred. No. 0.098; Mismatches 52; Indels 0; Gaps 0;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGATGGGGAATGCTCTGTGATACCTGTC 81





```
RESULT 55
AAK19029/c
ID AAK19029 standard; DNA; 292 BP.
XX
AC AAK19029;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 19020.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX
PT brains.
XX
PS Example 4; SEQ ID NO 19020; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX
CC probes which are derived from genomic sequences expressed in the human
XX
CC brain. They can be used to measure gene expression in brain cell samples,
XX
CC which may enable the diagnosis and improved treatment of nervous system
XX
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX
CC epilepsy and cancers. The present sequence is one of the probes of the
XX
CC invention
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACGAGTGTGCGGTGTGCGGAGCGGAGGAGCTCATCTGCTGTGACGGCTGCC 588
DB 140 ATATGGAATTCTGTGCGGTGTGCAAGATGTGTGGGAACCTGCTGTGATTACTGTTC 81
QY CTGCGGCTTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
DB 80 CTTCTTCTTAACCAATCACTGCTGAATCCCACTTCCAGAGATCCCAAGGTGAAT 21
QY 649 GGAAGTGTCTCAGCTG 664
DB 20 GGCTCTGTCCCGCTTG 5

RESULT 56
ABS44646/c
ID ABS44646 standard; DNA; 292 BP.
```

```
XX
AC ABS44646;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID NO 19636.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX
PT gene expression in human adult liver.
XX
PS Claim 4; SEQ ID NO 19636; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (i) for
XX
CC measuring human gene expression in a sample derived from human adult
XX
CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX
CC specification (or complements/ fragments). The probe hybridises at high
XX
CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX
CC (i) may be used for predicting, measuring and displaying gene expression
XX
CC in samples derived from human adult liver. The genes identified may be
XX
CC involved in genetic liver diseases such as cirrhosis,
XX
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
XX
CC liver single exon nucleic acid probes of the invention. Note: The
XX
CC sequence information for this patent does not appear in the printed
XX
CC specification but was obtained in electronic format directly from WIPO at
XX
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACGAGTGTGCGGTGTGCGGAGCGGAGGAGCTCATCTGCTGTGACGGCTGCC 588
DB 140 ATATGGAATTCTGTGCGGTGTGCAAGATGTGTGGGAACCTGCTGTGATTACTGTTC 81
QY CTGCGGCTTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
DB 80 CTTCTTCTTAACCAATCACTGCTGAATCCCACTTCCAGAGATCCCAAGGTGAAT 21
QY 649 GGAAGTGTCTCAGCTG 664
DB 20 GGCTCTGTCCCGCTTG 5

RESULT 57
```

ABSI9225/C  
ID ABSI9225 standard; DNA; 292 BP.  
XX  
AC  
XX ABSI9225;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human genome-derived single exon probe ORF from lung SEQ ID NO 19216.  
XX  
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Heremanky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-0060840B.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI, 2002-114183/15.  
XX  
PT Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
PS  
XX Claim 4; SEQ ID NO 19216; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression by a nucleic acid derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung

CC	cancer; chronic obstructive pulmonary disease (COPD), interstitial lung
CC	disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC	tuberosus sclerosis, Gaucher's disease, Niemann-Pick disease, Hemaneky-
CC	Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC	histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC	Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC	dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC	present sequence is a single exon probe open reading frame of the
CC	invention. Note: The sequence data for this patent did not form part of
CC	the printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 292 BP, 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
Query Match	3.4%; Score 52.8; DB 6; Length 292;
Best Local Similarity	61.8%; Pred. No. 0.098;
Matches	84; Conservative 0; Mismatches 52; Indels 0; Gaps 0
Dy	529 ATGAGACGAGTGTGCCCTGTGTGTCGGACGCGCGGAGACTATCTGTCTTGACGGTCCC 588
Dd	140 ATATGGAATTCTGTCTGCGGCTGTGCAGAAGATGTGGGGAACGTCTGTGTGATTAACCTGTC 81
Oy	589 CTCGGGCGCTTCCACCCTGCGCTGCGCTGCCCCCTCGCTCGGAGATPCCCCAGATGGGAACT 648
Dd	80 CTCTCTCTTACCACTCATCCAGTGCCTGAATCCCCCACTTCCAGAGATCCCAACGGTGAAT 21
Oy	649 GGAGGTGCTCCAGCTG 664
Dd	20 GGCTCTGTCCCGTTG 5
RESULT 58	
ID	AAI15873/c
ID	AAI15873 standard; DNA; 476 BP.
XX	
AC	AAI15873;
XX	
DT	12-OCT-2001 (first entry)
DE	Probe #5806 for gene expression analysis in human cervical cell sample.
XX	
KW	Probe; human; microarray; gene expression; cervical epithelial cell;
KW	cervical cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157278-A2.
PD	
PD	09-AUG-2001.
PE	
PF	30-JAN-2001; 2001WO-US000670.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
PI	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
WP	WI; 2001-488901/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human cervical epithelial cells.
XX	
ES	Claim 25; SEQ ID NO 5806; 487bp; English.
XX	
CC	The present invention relates to human single exon nucleic acid probes
CC	(SNMP). The present sequence is one such probe. The SNMPs are derived

CC from human HeLa cells. The SNPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging of  
CC diseases of the cervix, notably cervical cancer. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;  
SQ

Query Match 3.4%; Score 52.8; DB 4; Length 476;  
Best Local Similarity 61.8%; Pred. No. 0.1;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTCGGACGCGGGAGCTCATCTGCTGACGCTGCC 588  
DB 428 ATATGGAATTCTGTCGGTCTGCAAGATGATGGGAACTGCTGTGATACCTGTC 369  
QY 589 CTCGGGCGCTTCCACCTGGCGCTGTCCTCCCTCCGTCGGAGATCCCACTGAGGACCT 648  
DB 368 CTTCTTCTTACCACTTCACATCTGATTCCTCCCACTTCAGAGATCCCAAGCTGAAAT 309  
QY 649 GGAGGTGCTCCAGCTG 664  
DB 308 GGCTCTGTCCCGTTG 293

RESULT 59  
ABAS8175/C  
ID ABAS8175 standard; DNA; 476 BP.  
XX  
AC ABAS8175;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #6480.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000669.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human foetal liver.  
XX  
PS Claim 1; SEQ ID NO 6480; 639bp + Sequence Listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human foetal liver. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;  
SQ

Query Match 3.4%; Score 52.8; DB 4; Length 476;  
Best Local Similarity 61.8%; Pred. No. 0.1;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTCGGACGCGGGAGCTCATCTGCTGACGCTGCC 588  
DB 428 ATATGGAATTCTGTCGGTCTGCAAGATGATGGGAACTGCTGTGATACCTGTC 369  
QY 589 CTCGGGCGCTTCCACCTGGCGCTGTCCTCCCTCCGTCGGAGATCCCACTGAGGACCT 648  
DB 368 CTTCTTCTTACCACTTCACATCTGATTCCTCCCACTTCAGAGATCCCAAGCTGAAAT 309  
QY 649 GGAGGTGCTCCAGCTG 664  
DB 308 GGCTCTGTCCCGTTG 293

RESULT 60  
AA137782/C  
ID AA137782 standard; DNA; 476 BP.  
XX  
AC AA137782;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Probe #6468 used to measure gene expression in human placenta sample.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488997/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.  
XX  
PS Claim 25; SEQ ID NO 6468; 654bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SNP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders

XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;  
SQ

Query Match 3.4%; Score 52.8; DB 4; Length 476;  
Best Local Similarity 61.8%; Pred. No. 0.1;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy		529	ATGAGACGAGTGTCCCGTGTCGGAGACGGGGGAGCTCATCTGTGTGAAGGCTGCC	588
Dd		428	ATATGGAAATTCTGTGGGCTCTCAAGAGATGTGGGGAATGCTCTGCTGTGTACTGTGC	369
Oy		589	CTCGGGCCTTCCACCTGTGCTGCTCCCTCCGCTCCGGAGATCCCCAAGTGGACCT	648
Dd		368	CTTCTCTTCCACCAATCCACTCTGTAATCCCCCACTTCAAGATCCCCAACGGTAAT	309
Oy		649	GGAGGTGCTCCAGCTG	664
Dd		308	GAGCTGTCTCCGTTG	293
 RESULT 61 ABA27360/C				
ID			ABA27360 standard; DNA; 476 BP.	
AC			ABA27360;	
XX				
DT			23-JAN-2002 (first entry)	
XX				
DE			Probe #5826 for gene expression analysis in human heart cell sample.	
KX			Human; gene expression; heart; microarray; vascular system; probe;	
KW			cardiovascular disease; hypertension; cardiac arrhythmia;	
XX			congenital heart disease; ss.	
OS			Homo sapiens.	
XX				
PN			WO200157274-A2.	
PD				
XX				
XX			09-AUG-2001.	
PF				
XX			30-JAN-2001; 2001WO-US000666.	
PR				
XX			04-FEB-2000; 2000US-0180312P.	
PR			26-MAY-2000; 2000US-0207456P.	
XX			30-JUN-2000; 2000US-00608408.	
PR			03-AUG-2000; 2000US-00632366.	
XX			21-SEP-2000; 2000US-0234687P.	
PR			27-SEP-2000; 2000US-0236359P.	
XX			04-OCT-2000; 2000GB-00024263.	
PA			(MOLE-) MOLECULAR DYNAMICS INC.	
XX				
XX				
P1			Penn SG, Hanzel DK, Chen W, Rank DR;	
XX				
DR			WPI; 2001-488699/53.	
FT				
XX				
ST			Single exon nucleic acid probes for analyzing gene expression in human hearts.	
PS				
XS			Claim 1; SEQ ID NO 5826; 530bp; English.	
XX				
CC			The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosting diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
CC				
XX				
SO			Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;	
Query_Match		3.4%;	Score 52.8; DB 4; Length 476;	
Best Local Similarity		61.8%;	Fred. No. 0.1;	
Matches		84; Conservative	0; Mismatches	52; Indels
			0; Gaps	0

Oy		529	ATGAGACGAGAGTGCCTCCGTGTGTCGGAGACGGCGGGAGACTCATCTGCTGTGAAGGCTGCC	588
Dd		428	ATATGAAATTCTGTGGGCTCTTCGAAGAAGATGATGGGAACTGCTCTGTGTGTAACCTGTC	369
Oy		589	CTCGGACCCTTCCACCTTGACCTGCTGCTGTCCCTCCTCGCTCGGAGATCCCCAGTGGACCT	648
Dd		368	CTTCTTCTCAACACATTCACACTGCTCGAATATCCCCCACTTCCAGATGCCAACGGTGAAT	309
Oy		649	GGAGGTGCTCCAGCTG	664
Dd		308	GGCTCTGTCCCGCTTG	293
 RESULT 62 AAK31915/c ID AAK31915 standard; DNA; 476 BP. XX AC AAK31915; XX DT 06-NOV-2001 (first entry) XX DE Human bone marrow expressed single exon probe SEQ ID NO: 6472. XX KW Human; bone marrow expressed exon; gene expression analysis; probe; KW microarray; cancer; leukaemia; lymphoma; myeloma; ss. XX OS Homo sapiens. XX PN WO200157276-A2. PN PD 09-AUG-2001. PD PF 30-JAN-2001; 2001WO-US000668. XX PR 04-FEB-2000; 2000US-0180312P. PR 26-MAY-2000; 2000US-0207456P. PR 30-JUN-2000; 2000US-00608408. PR 03-AUG-2000; 2000US-00632366. PR 21-SEP-2000; 2000US-0234687P. PR 27-SEP-2000; 2000US-0236359P. PR 04-OCT-2000; 2000GB-00024263. XX PA (MOLE-) MOLECULAR DYNAMICS INC. XX PI Penn SG, Hanzel DK, Chen W, Rank DR; XX DR WPI; 2001-488900/53. XX PT Human genome-derived single exon nucleic acid probes useful for analyzing PT gene expression in human bone marrow. PT PS Example 4; SEQ ID NO 6472; 658bp + Sequence Listing; English. XX CC The present invention provides a number of single exon nucleic acid CC probes which are derived from genomic sequences expressed in the human CC bone marrow. They can be used to measure gene expression in bone marrow CC samples, which may enable the improved diagnosis and treatment of cancers CC such as lymphoma, leukaemia and myeloma. The present sequence is one of CC the probes of the invention XX SQ Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;				
Oy	Query Match	3 4%;	Score 52.8;	DB 4; Length 476;
Oy	Best Local Similarity	61.8%;	Pred. No. 0.1;	52; Indels 0; Gaps 0
Oy	Matches	84; Conservative	0; Mismatches	
Dd		529	ATGAGACGAGAGTGCCTCCGTGTGTCGGAGACGGCGGGAGACTCATCTGCTGTGAAGGCTGCC	588
Dd		428	ATATGAAATTCTGTGGGCTCTTCGAAGAAGATGATGGGAACTGCTCTGTGTGTAACCTGTC	369
Oy		589	CTCGGACCCTTCCACCTTGACCTGCTGCTGTCCCTCCTCGCTCGGAGATCCCCAGTGGACCT	648
Dd		368	CTTCTTCTCAACACATTCACACTGCTCGAATATCCCCCACTTCCAGATGCCAACGGTGAAT	309
Oy		649	GGAGGTGCTCCAGCTG	648
Dd		368	CTTCTTCTCAACACATTCACACTGCTCGAATATCCCCCACTTCCAGATGCCAACGGTGAAT	309

QY 649 GGAGGCTCCAGCTG 664  
DB 308 GGCTCTGCCCGTTG 293

RESULT 63  
AAK06255/c  
ID AAK06255 standard; DNA; 476 BP.

AC AAK06255;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 6246.

KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
KW ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000667.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains.

PS Example 4; SEQ ID NO 6246; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention

XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

QY Query Match 3.4%; Score 52.8; DB 4; Length 476;  
Best Local Similarity 61.8%; Pred. No. 0.1;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCTGTGTCGGAGCGGAGGAGCTCATCTGTGACGCTGCC 588

DB 428 ATATGGAATTCTGTGCGGCTCTGCAAGATGTGGGAATGCTCTGTGTGATCTGTTC 369

QY 589 CTCGGGCTTCACCTGCGCTGTGCTCCCTCGCTCCGGAGATCCCAAGTGGAGCT 648

DB 368 CTTCTTCTTACCACTGCTGCTGAATCCCACTTCAGATGCCCAAGTGAAT 309

QY 649 GGAGGCTCCAGCTG 664

DB 308 GGCTCTGCCCGTTG 293

RESULT 64  
ABS31605/c  
ID ABS31605 standard; DNA; 476 BP.

AC ABS31605;

DT 25-FEB-2003 (first entry)

DE Human liver single exon probe, SEQ ID NO 6595.

KW Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.

OS Homo sapiens.

PN WO200157273-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000664.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-48898/53.

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.

PS Claim 1; SEQ ID NO 6595; 658bp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (1) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 1109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (1) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
CC liver single exon nucleic acid probes of the invention. Note: The  
CC sequence information for this patent does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

QY Query Match 3.4%; Score 52.8; DB 4; Length 476;  
Best Local Similarity 61.8%; Pred. No. 0.1;  
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCTGTGTCGGAGCGGAGGAGCTCATCTGTGACGCTGCC 588

DB 428 ATATGGAATTCTGTGCGGCTCTGCAAGATGTGGGAATGCTCTGTGTGATCTGTTC 369

QY 589 CTCGGGCTTCACCTGCGCTGTGCTCCCTCGCTCCGGAGATCCCAAGTGGAGCT 648

DB 368 CTTCTTCTTACCACTGCTGCTGAATCCCACTTCAGATGCCCAAGTGAAT 309

QY 649 GGAGGCTCCAGCTG 664

DB 308 GGCTCTGCCCGTTG 293

DB 308 GGCTCTGTCCTG 293

RESULT 65  
ABSO6677/c  
ID ABSO6677 strand; DNA; 476 BP.

AC ABSO6677;  
XX  
XX  
XX 19-AUG-2002 (first entry)  
DE Human genome-derived single exon probe from lung SEQ ID NO 6668.

XX Human; de; single exon probe; asthma; lung cancer; COPD; ILD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.

XX Homo sapiens.  
XX  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI, 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.

XX  
XX Claim 1, SEQ ID NO 6668; 634bp; English.

XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, primary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a single exon probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

XX  
XX SQ Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

XX  
XX Query Match 3.4%; Score 52.8; DB 6; Length 476;  
XX Best Local Similarity 61.8%; Pred. No. 0.1;  
XX Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

XX  
XX 529 ATGAGACGAGTGTCCGTCGAGCGGCGGAGCTCATCTGTGTGACGGCTGCC 588  
XX |||||  
XX 428 ATATGGAATTCGTGGGTCTGCAGAGATGGGGAACTGCTGTGTGATCTGTC 369  
XX |||||  
XX 589 CTCGGACCTTCCACCTGGCTGCTGTCCTCCCTCGCTCCGGAGATCCCACTGGACCT 648  
XX |||||  
XX 368 CTTCTTCCATCACATCCATCCATGTAATCCCTCCATCCAGATGCCAAGTGTAAT 309  
XX |||||  
XX 649 GGAGGTGCTCCAGCTG 664  
XX |||||  
XX DB 308 GGCTCTGTCCTG 293

XX  
XX RESULT 66  
XX ABL29649  
XX ID ABL29649 strand; DNA; 6025 BP.

XX  
XX ABL29649;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 40420.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; de.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX  
XX WPI, 2001-656660/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.

XX  
XX Claim 1, SEQ ID NO 40420, 21pp + Sequence Listing; English.

XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-  
CC ABB82072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 6025 BP; 1591 A; 1529 C; 1756 G; 1149 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 6025;  
Best Local Similarity 54.7%; Pred. No. 0.15;

Matches 105; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
QY 522 CAGAAATGAGACGAGTGTGCTGTGTCGAGACGGGGAGCTCATCTGCTGTAC 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1093 CATGAGCACACGACTACTGCGAGGTGTCCAGCAGAAGAGTGAGATCCTGTGCGAC 1152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 582 GGGTGGCCCTGGGGCTTCCAGCTGCTGCTGCTCCCTCCGCTCCGGAGATCCCACT 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1153 ACCTGCTCTGGGACATATCATCTTGTGTCTGTGAGCCAGAACTCGATTAACCGCCAGAG 1212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 642 GGGACCTGAGGTGTCTCCAGCTGCTGCGACGACAGTCCAGAGGTGCGACCCCGGGCA 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1213 GGCAGTGTGCTGTGCTCGCACTGCGAGGCTGAACGAGGTGCTGTGAGGAAGAGACGAT 1272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 702 GAGAGACCCCGG 713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1273 GATGAGCACACG 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 67

AB129648/c  
ID ABL29648 standard; DNA; 7303 BP.

XX ABL29648;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 40417.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmacetical; gene; de.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.

XX Claim 1; SEQ ID NO 40417; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-  
CC ABB82072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 7303 BP; 1509 A; 2002 C; 1786 G; 2006 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 7303;  
Best Local Similarity 54.7%; Pred. No. 0.15;

Matches 105; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```
QY 522 CAGAAATGAGACGAGTGTGCTGTGTCGAGACGGGGAGCTCATCTGCTGTAC 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5509 CATGAGCACACGACTACTGCGAGGTGTCCAGCAGAAGAGTGAGATCCTGTGCGAC 5450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 582 GGGTGGCCCTGGGGCTTCCAGCTGCTGCTGCTCCCTCCGCTCCGGAGATCCCACT 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5449 ACCTGCTCTGGGACATATCATCTTGTGTGCTGTGAGCCAGAACTCGATTAACCGCCAGAG 5390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 642 GGGACCTGAGGTGTCTCCAGCTGCTGCGACGACAAATCCAGAGTGCAGCCCGGGCA 701
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5389 GGCAGTGTGCTGTGCTCGCACTGCGAGGCTGACGAGGTGCTGTGAGGAAGAGACGAT 5330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 702 GAGAGACCCCGG 713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5329 GATGAGCACACG 5318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 68

AB099469  
ID AB099469 standard; cDNA; 4653 BP.

XX AB099469;

XX 25-FEB-2003 (first entry)

XX Human coding sequence SEQ ID 202.

XX Human; expressed sequence tag; EST; chromosome 1;  
XX haematopoietic disorder; central nervous system disease; viral infection;  
XX peripheral nervous system disease; non-healing wound; infectious disease;  
XX immune deficiency; immune disorder; bacterial infection; allergy; cancer;  
XX fungal infection; autoimmune disorder; coagulation disorder; noctropic;  
XX anti-allergic; anti-inflammatory; immunosuppressive; neuroprotective;  
XX cytostatic; haemostatic; virucide; antibacterial; fungicide;  
XX immunostimulant; cerebroprotective; gene therapy; gene; ss.

XX Homo sapiens.

XX WO200259260-A2.

XX 01-AUG-2002.

XX 16-NOV-2001; 2001WO-US042950.

XX 17-NOV-2000; 2000US-00714936.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;

XX Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-590824/63.

XX N-PSDB; ABP64883.

XX New isolated polynucleotide, useful in research, diagnostic or  
XX therapeutic methods, e.g. preventing or treating disorders involving  
XX aberrant protein expression or biological activity.

XX Claim 1; SEQ ID NO 202; 394bp; English.

CC The present invention relates to novel human coding sequences (AB099268-



CC AB099608) and proteins (ABP64682-ABP65022). The sequences are useful in  
CC therapeutic, diagnostic and research methods. The polynucleotides may be  
CC used in the field of molecular biology as hybridisation probes, primers  
CC for PCR, for chromosome and gene mapping, for the recombinant production  
CC of protein, or in generation of anti-sense DNA or RNA. The  
CC polynucleotides are useful in diagnostics as expressed gene tags  
CC (ESTs) for identifying expressed genes or for physical mapping of the  
CC human genome. The proteins may be used as molecular weight markers, or as  
CC nutritional sources or supplements. The proteins may be used to maintain  
CC and expand cell population in a totipotent or pluripotent state  
CC useful for re-engineering damaged or diseased tissues, transplantation,  
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
CC polynucleotides and proteins are useful for preventing, treating or  
CC ameliorating disorders involving aberrant protein expression or  
CC biological activity, e.g. haematopoietic disorders, central/peripheral  
CC nervous system diseases, mechanical and traumatic disorders, non-healing  
CC wounds, immune deficiencies and disorders, infectious diseases caused by  
CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
CC reactions and conditions, coagulation disorders, or cancer. The  
CC polynucleotide sequences of the invention were assembled from ESTs  
CC isolated mainly by sequencing by hybridisation, and in some cases,  
CC sequences obtained from one or more public databases. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 4653 BP; 1097 A; 1368 C; 1416 G; 772 T; 0 U; 0 Other;

XX Query Match 3.4%; Score 52.6; DB 6; Length 4653;

XX Best Local Similarity 66.1%; Pred. No. 0.16; Mismatches 39; Indels 0; Gaps 0;

XX Matches 76; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 533 GGACGAGTGTGCGGTGTGCGGAGCGGAGCTCATCTGTGTGACGGCTGCGCCTG 592  
DB 8 GGAGTCTCTCCCGGTGTGCGAGCGGTTGCGAGTCTCTGTGCGAGCGCTGCGCCTG 67

OY 593 GGGCTTCACAGCTGCGCTGCTGCTCCCTCGCGGAGATCCCAAGTGGAGC 647  
DB 68 CTCCTACCACTGATGCTGCTCAACCGCGCTGCGGAGATCCCAAGTGGAGC 122

RESULT 69

ABV44220  
ID ABV44220 standard; cDNA; 539 BP.

XX ABV44220;

XX 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 44211.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

OS WO200160860-A2.

PN 23-AUG-2001.

PD 20-FEB-2001; 2001WO-US005171.

PF 17-FEB-2000; 2000US-018319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PA Schlegel R, Endege WO, Monahan JB;

XX PT

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8783; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 539 BP; 125 A; 148 C; 150 G; 115 T; 0 U; 1 Other;

XX Query Match 3.4%; Score 52.2; DB 5; Length 539;

XX Best Local Similarity 66.4%; Pred. No. 0.14; Mismatches 38; Indels 0; Gaps 0;

XX Matches 75; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 553 GGGACGGCGGAGCTCATCTGTGACGCGCTGCGGAGCTTCACCTGCGCTGCC 612  
DB 50 GGGACGGCGGAGGCTCTGTGTGACGCGGTGATCTCTTACCATTCATTGTC 109

OY 613 TGTCCCTCTCCGCTCCGGAGATCCCAAGTGGAGCTGAGGTGCTTCAGCTGC 665  
DB 110 TTAACCTCTCCCTGCTGATTCCTCCCAATGGAATGCTGTGCTCCCGATGC 162

RESULT 70

ABV29094  
ID ABV29094 standard; cDNA; 539 BP.

XX ABV29094;

XX 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 29085.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

OS WO200160860-A2.

PN 23-AUG-2001.

PD 20-FEB-2001; 2001WO-US005171.

PF 17-FEB-2000; 2000US-018319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PA Schlegel R, Endege WO, Monahan JB;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.



KW immunogen; cancer; actinic keratosis; arteriosclerosis; atherosclerosis;  
 KW buruli's; cirrhosis; hepatitis; psoriasis; AIDS; rheumatoid arthritis;  
 KW adult respiratory distress syndrome; Addison's disease; allergy; anaemia;  
 KW asthma; osteoporosis; autoimmune; haemolytic anaemia; scleroderma;  
 KW autoimmune thyroiditis; Crohn's disease; atopic dermatitis;  
 KW diabetes mellitus; Graves' disease; glomerulonephritis;  
 KW systemic lupus erythematosus; systemic sclerosis; ulcerative colitis;  
 KW hemodialysis; uveitis; trauma; Alzheimer's; Pick disease;  
 KW Parkinson's disease; amyotrophic lateral sclerosis; epilepsy; stroke;  
 KW Huntington's disease; multiple sclerosis; dementia;  
 KW extrapyramidal disorder; motor neuron disorder; central nervous system;  
 KW neuromuscular disorder; metabolic; endocrine; toxic myopathy;  
 KW periodic paralysis; mental disorder; human; gene; ds.  
 KM Homo sapiens.  
 OS  
 XX  
 XX W0200296951-A1.  
 XX  
 XX  
 XX 05-DEC-2002.  
 XX  
 XX  
 XX  
 XX 24-MAY-2002; 2002MO-US016676.  
 XX  
 XX 25-MAY-2001; 2001US-0293723P.  
 PR 01-JUN-2001; 2001US-0293257P.  
 PR 08-JUN-2001; 2001US-0297220P.  
 PR 21-JUN-2001; 2001US-0300526P.  
 PR 29-JUN-2001; 2001US-0301874P.  
 PR 22-FEB-2002; 2002US-0359413P.  
 XX  
 XX  
 XX (INCV-) INCVTE GENOMICS INC.  
 PA  
 XX  
 XX Tang TY, Yue H, Baughn MR, Duggan BM, Warren BA, Bandman O;  
 PI Tranekun TM, Burford N, Sanjanwala B, Becha SD, Yao MG, Yang J;  
 PI Rhan US, Hafalia AJA, Griffin JA, Swarnakar A, Elliott VS;  
 PI Rection SA, Khan PA, Lee EA, Yue H, Lu DAM, Walla NK, Thangavelu K;  
 PI Avizien CS, Xu Y, Ison CH, Huang J, Ding L, Honchell CD, Mason PM,  
 PI Borowsky ML, Emerling BM, Peterson DP, Lu Y, Ramkumar J, Mason PM,  
 PI Zabarjadian Y, Azimzai Y, Stuve LL, Kamigaki LL, Barroso I, Lee S;  
 PI Kable AE;  
 XX  
 XX WPI; 2003-140448/13.  
 DR P-PSDB; AAO262249.  
 DR  
 PT Novel molecules for disease detection and treatment and polynucleotide  
 PT encoding them useful for diagnosing, preventing or treating cell  
 PT proliferative, autoimmune/inflammatory, neurological and developmental  
 PT disorders.  
 PS  
 XX  
 XX Claim 121; Page 248; 260pp; English.  
 CC The invention relates to an isolated polypeptide chosen from molecules  
 CC for disease detection and treatment (MDDT), comprising a one of 39 114-  
 CC 1250 residue amino acid sequences, given in the specification, or a  
 CC biologically active or immunogenic fragment of the isolated polypeptide.  
 CC The isolated polypeptide is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist of the isolated polypeptide.  
 CC The isolated polypeptide is also useful as an immunogen for preparing  
 CC polyclonal or monoclonal antibodies by hybridoma technology. The isolated  
 CC polypeptide and its encoding polynucleotide are useful for diagnosis,  
 CC treatment and prevention of cancer, actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, buruli's, cirrhosis, hepatitis, psoriasis, AIDS, adult  
 CC respiratory distress syndrome, Addison's disease, allergies, anaemia,  
 CC asthma, atherosclerosis, osteoporosis, autoimmune haemolytic anaemia,  
 CC autoimmune thyroiditis, Crohn's disease, atopic dermatitis, diabetes  
 CC mellitus, Graves' disease, glomerulonephritis, rheumatoid arthritis,  
 CC scleroderma, systemic lupus erythematosus, systemic sclerosis, ulcerative  
 CC colitis, haemodialysis, uveitis, viral, bacterial, fungal, parasitic,  
 CC protozoal, helminthic infections, trauma, Alzheimer's and Pick disease,  
 CC Parkinson disease, amyotrophic lateral sclerosis, epilepsy, stroke,  
 CC Huntington's disease, multiple sclerosis, dementia, and other  
 CC extrapyramidal disorder, motor neuron disorder, and other developmental  
 CC disorders of the central nervous system, neuromuscular disorders,  
 CC metabolic, endocrine and toxic myopathies, periodic paralysis, mental

disorders including mood, anxiety and schizophrenia disorders, anaemia, renal tubular acidosis, epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and cataract. This polynucleotide sequence represents the DNA encoding a human MDRP protein relating to the invention

SQ Sequence 1965 BP; 439 A; 664 C; 557 G; 305 T; 0 U; 0 Other;

Query Match            3.4%; Score 52; DB 10; Length 1965;  
Best Local Similarity   57.3%; Pred.No. 0.19;  
Matches     94; Conservative   0; Mismatches   70; Indels        0; Gaps        0

OY    529 ATGAGCAGACGTGGCCCGTCTGTCTGCAGACGGGGGAGCTACTTCGTGTGAACGCTGCC 588  
| | | | | | | | | | | | | | | | | | | | | |  
DB    1085 ACATGAGCACACTGTGCCCCCTTGCAAGCAGGGGCCAACCTGCACCCTTGCAGCATTTGCC 114  
| | | | | | | | | | | | | | | | | | | | | |  
OY    589 CTGGGGCCCTTCCACACTGTGCGCTGCTCTGCCCTCCGCTCCGGAGAGATTCCCAGTGGACCT 648  
| | | | | | | | | | | | | | | | | | | | | |  
DB    1145 CGGGGGCCCTTACCACCTCACTGACTGTCTGTGAGCCGCCCTTAAGAAGCGCGCCCAAGGGCGTGT 120  
| | | | | | | | | | | | | | | | | | | | | |  
OY    649 GGAGTGTCTCACTGCTGCTGTCAGGACAAGTCACAGAAGTGTGAG 692  
| | | | | | | | | | | | | | | | | | | | | |  
DB    1205 GGATGTGCCCCAGGTGCAGCAGAAAGCCCTTAAGAAGACGAG 1248  
| | | | | | | | | | | | | | | | | | | | | |

RESULT 73  
ADM02221  
ID ADM02221 standard; cDNA; 2651 BP.

XX    ADM02221;  
XX  
XX  
XX    20-MAY-2004 (first entry)  
DT  
DE Human CDNA of the invention SEQ ID NO:906.  
XX  
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.  
OS Homo sapiens.  
XX  
PN EP1347046-A1.  
XX  
PD 24-SEP-2003.  
XX  
PF 12-APR-2002; 2002EP-00008400.  
XX  
PR 22-MAR-2002; 2002JP-00137785.  
XX  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX    Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
PI Yamamoto J, Isono Y, Hiro Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
PI Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuno Y;  
XX WPI; 2003-723558/69.  
DR P-Psdb; ADM04664.  
XX

Pt New polynucleotides and polypeptides are useful in gene therapy, for  
Pt developing a diagnostic marker or medicines for regulating their  
expression and activity, or as a target of gene therapy.

XX Claim 1; SEQ ID NO 906; 305bp; English.  
ES  
XX The invention relates to a novel human polynucleotide and the encoded  
polypeptide. A polynucleotide of the invention ADM06202-ADM06773 is useful  
as a primer for synthesizing the polynucleotide or as a probe for  
detecting the polynucleotide. The polynucleotides ADM0331c-ADM03758 are  
useful in gene therapy, for developing a diagnostic marker or medicines  
for regulating their expression and activity, or as a target of gene  
therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
are useful as pharmaceutical agents. The present sequence represents a  
cDNA sequence of the invention.

SX Sequence 2651 BP; 577 A; 730 C; 779 G; 565 T; 0 U; 0 Other;



DR MPI, 2003-450961/43.  
DR P-PSDB, ADB64354.

XX New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.

XX Claim 1; Page; 222pp; English.

PS The invention discloses a polynucleotide comprising a sequence selected  
XX from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumors). The CDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a CDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.

XX SQ Sequence 3584 BP; 770 A; 1089 C; 1030 G; 695 T; 0 U; 0 Other;

Query Match 3.4%; Score 52; DB 10; Length 3584;  
Best Local Similarity 57.3%; Pred. No. 0.21; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 70; Indels 0; Gaps 0;

OY 529 ATGAGAGCAGAGTGTGCGGTGTGCGGAGCGGCGGAGAGTCATCTGTGACGGCTGCC 588  
DB 1118 ACATATAGACATGTGTGCGGCTGCGAGGAGGCGGCAACCTGAGCCCTGGGCACTGCC 1177  
OY 589 CTCGGGCTTTCCACCTGCGGCTGCTCCCTCCGCTCCGCGAGATCCCAAGTGGAACT 648  
DB 1178 CGGGGGCTTACCACTGAGTGTGCTGAGCGCGCCCTCAAGAGCGGCGCCCAAGGGCGTGT 1237  
OY 649 GGAAGTGTCTCAGCTGCGCTGCGGCAACAGTCCAGGAGGTGAG 692  
DB 1238 GGGTGTGCCCCAGGTGCGGAGGAGGCTTTAAAGAAAGCGAG 1281

RESULT 76  
ADQ24519  
ID ADQ24519 standard; DNA; 3710 BP.

XX ADQ24519;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 7339.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
KW ds.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnick A;

XX MPI, 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

XX Example 2; SEQ ID NO 7339; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC DNA of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 3710 BP; 784 A; 1129 C; 1032 G; 698 T; 0 U; 67 Other;

Query Match 3.4%; Score 52; DB 12; Length 3710;  
Best Local Similarity 57.3%; Pred. No. 0.21; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 70; Indels 0; Gaps 0;

OY 529 ATGAGAGCAGAGTGTGCGGTGTGCGGAGCGGCGGAGAGTCATCTGTGACGGCTGCC 588  
DB 1203 ACATATAGACATGTGTGCGGCTGCGAGGAGGCGGCAACCTGAGCCCTGGGCACTGCC 1262  
OY 589 CTCGGGCTTTCCACCTGCGGCTGCTCCCTCCGCTCCGCGAGATCCCAAGTGGAACT 648  
DB 1263 CGGGGGCTTACCACTGAGTGTGCTGAGCGCGCCCTCAAGAGCGGCGCCCAAGGGCGTGT 1322

OY 649 GGAAGTGTCTCAGCTGCGCTGCGGCAACAGTCCAGGAGGTGAG 692  
DB 1323 GGGTGTGCCCCAGGTGCGGAGGAGGCTTTAAAGAAAGCGAG 1366

RESULT 77  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 DR WPI, 2003-175290/17.  
 XX  
 XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 XX Claim 27, SEQ ID NO 5263, 899pp; English.  
 PS  
 XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 XX Sequence 2000 BP, 336 A, 265 C, 284 G, 363 T, 0 U, 752 Other;  
 SQ  
 Query Match 3.3%; Score 50.6; DB 8; Length 2000;  
 Best Local Similarity 10.9%; Pred. No. 0.38;  
 Matches 69; Conservative 293; Mismatches 262; Indels 7; Gaps 3;  
 QY 599 CCACCTGGCTGCTGCTCCCTCCCTCCGCTCCGAGATCCCACTGAGACTGAGTGCCTC 658  
 DB 4 CSAGWRSRCSWMGRBRMYMAGMWCARWGSSRMSRMSKYSRKCSCGCKMTTRK 63  
 QY 659 CAGCTGCTCGCAGCAAGTCCAGAGAGTGCAGCCCGGAGAGAGAGAGAGAGAGAGAGAG 718  
 DB 64 SKWYSASASGRTSGSWSSGSYSKGKMKRKYRSKMRGRGRMRSRMRMRGRYRCA 123  
 QY 719 GGAGCACCCTGGAGAGACCCCGCTCCCGGGGCTTAAAGTCGCGAGAGAGAGAGTAAG 778  
 DB 124 RGRMAGSGSRMWSGSRMSYMMWCYARCGSCIKRKSXGSGSKTCRGRAGRSGMSWG 183  
 QY 779 AGGTCCACTGGGGAACCCCTAGCCGCAATGACACACTCTTGTCT-ACAAGCACTGTC 837  
 DB 184 AAKYS--GSMKMKMMWSGCRSGCGRRSAYSRYGTSRKTYTKMTYASASRCKMAVMT 241  
 QY 838 CGGCTCGGCTTGTGACGCCCGCTCCAGTGTGACTCTCGGCTTCGACCCCTCAG 897  
 DB 242 TSTSWACSSYTWCRSRRRMMWMMKMRKMSRSYGMYSYKMMCTAYKKSYSRWCYM 301  
 QY 898 TGTGTGTGGTCTCTGAGAGTCAAGAACCTGCTCTGTCGCGCTTCCGGGGTGTGCG 957  
 DB 302 YGGGGRGATRYWGRYMRMAMMYKMTMYNGYKMKRGMWAGRMWMSMCRMSACRY 361  
 QY 958 GAGATGTAAGACGAGTGTGCGGTGTACTCACTGCGCCCTGCTTCACTGCGGTGCC 1017  
 DB 362 MWRMRMRTRRRRMAKKSRTSRSKRKRKCMRKRYKMRGYSRMSCKRARMMKRCRSG 421  
 QY 1018 ACTTCCAGCCGGCACTCCCGCGCGGAGACGGGCTCGGTGCAATCTGTCTGAGAG 1077  
 DB 422 RMMKMGCRGCMTCRMMSYGMWRKMSWKRMASIKYKMMRSKKSRTTMMGKTRGGM 481  
 QY 1078 AGGTGACCCCAAGCCCTGTGAGAGGGGTCTGCGCCCAAGC-----CCGCGCGCTGCGC 1133  
 DB 482 MGTMRCTRKRKRSKMGMRKRRRRRMRGMYMRMYMSARYMRYCARKKYISARAKAR 541  
 QY 1134 CCTGGGCTCCCAAGATACACTGCACTCAAGACCGGCTTCTGACAGGATGACCTG 1193  
 DB 542 CWYRGKYWAGMMWRKRYRMYYKMMYKRYKYSCKMSYVASKMSYASCKMSARKAGAM 601  
 QY 1194 GAGTCCCTTCTGAGCGAGACACCTTCGATG 1224  
 DB 602 CKRSKMSAWSKMSRSRKCRKCAKSKRSAR 632

RESULT 78  
 ADJ37277  
 ID ADJ37277 standard; cDNA, 2226 BP.  
 XX  
 XX ADJ37277;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 XX Human nucleic-acid associated protein NAAp-31 encoding cDNA SEQ ID NO:66.  
 DE  
 XX human; nucleic-acid associated protein; NAAp-31; NAAp; antiparkinsonian;  
 KW neurotrophic; neuroprotective; muscular; neuroleptic; antidiabetic;  
 KW antihypoid; cytostatic; dermatological; antiinflammatory;  
 KW immunosuppressive; antiallergic; nephrotropic; virucide; antibacterial;  
 KW fungicide; antiparasitic; protozoacide; antihelminthic; hepatotropic;  
 KW antileukosclerotic; cardiovascular; gastrointestinal; Parkinson's disease;  
 KW gene therapy; neurodegenerative disorder; Parkinson's disease;  
 KW Alzheimer's disease; muscular disorder; myotonic dystrophy; catatonia;  
 KW endocrine disorder; diabetes; Grave's disease; cancer; leukaemia;  
 KW cervical cancer; breast cancer; immunological disorder; scleroderma;  
 KW systemic lupus erythematosus; allergy; gastrointestinal disorder;  
 KW Crohn's disease; renal disorder; Goodpasture's syndrome; infection;  
 KW cardiovascular disorder; atherosclerosis; hepatic diseases; cirrhosis;  
 KW gene; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 90..1832  
 FT /\*tag= a  
 FT /\*product= "nucleic-acid associated protein NAAp-31"  
 FT  
 PN MO2004011604-A2.  
 XX  
 PD 05-FEB-2004.  
 XX  
 XX 23-JUL-2003; 2003WO-US023245.  
 PF  
 XX 26-JUL-2002; 2002US-0398907P.  
 PR 30-AUG-2002; 2002US-0407068P.  
 PR 26-SEP-2002; 2002US-0414139P.  
 PR 05-NOV-2002; 2002US-0424094P.  
 PR 17-JAN-2003; 2003US-0440912P.  
 PR 24-JAN-2003; 2003US-0442419P.  
 PR  
 XX  
 XX (INCY-) INCYTE CORP.  
 PA  
 XX Mason PM, Swarnakar A, Jiang X, Jackson AA, Kable AE, Tang YT,  
 PI Forsythe JD, Elliott VS, Lee SY, Lee BA, Ison CH, Hafalia AJA,  
 PI Khare R, Margue JP, Becha SD, Bulloch SA, Blake JF, Gandhi AR,  
 PI Griffin JA, Lee S, Yue H, Yang YG, Sprague WW, Baughn MR, Wang JT,  
 PI Gera M, Gietzen KJ, Nguyen DB, Lu DM;  
 XX  
 XX WPI, 2004-143837/14.  
 DR P-PSDB; ADJ37242.  
 XX  
 PT New nucleic acid-associated proteins (NAAp) polypeptides, useful for  
 PT diagnosing, preventing and treating disorders with abnormal activity of  
 PT NAAp, e.g. neurological, immunological, cardiovascular disorders and  
 PT cancer.  
 XX  
 PS Claim 5, SEQ ID NO 66; 290pp; English.  
 XX  
 XX The present sequence encodes a human nucleic-acid associated protein  
 CC designated NAAp-31. NAAp sequences have antiparkinsonian, neurotropic,  
 CC neuroprotective, muscular, neuroleptic, antidiabetic, antihypoid,  
 CC cytostatic, dermatological, antiinflammatory, immunosuppressive,  
 CC antiallergic, nephrotropic, virucide, antibacterial, fungicide,  
 CC antiparasitic, protozoacide, antihelminthic, antileukosclerotic,  
 CC cardiovascular, gastrointestinal and hepatotropic activities, and can be  
 CC used in gene therapy. The NAAp polypeptide or its fragments, and the  
 CC polynucleotide encoding the polypeptide are useful in diagnosing,

CC preventing, and treating disorders associated with an abnormal expression  
CC or activity of MAP (nucleic acid-associated proteins), such as  
CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's  
CC disease), muscular disorders (e.g. myotonic dystrophy, catatonias),  
CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.  
CC leukemia, cervical or breast cancers), immunological disorders (e.g.  
CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal  
CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's  
CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,  
CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis),  
CC or hepatic diseases (e.g. cirrhosis). NAAp or its fragments may also be  
CC used in screening for compounds that specifically bind to and modulate  
CC the activity of NAAp. The polynucleotides can be used to create humanised  
CC animals or transgenic animals to model human disease.

XX  
SQ Sequence 2226 BP; 702 A; 459 C; 583 G; 482 T; 0 U; 0 Other;

Query Match 3.2%; Score 48.8; DB 12; Length 2226;  
Best Local Similarity 65.7%; Pred. No. 0.93; Mismatches 0; Gaps 0;  
Matches 71; Conservative 0; Indels 0;

QY 515 CCTGTGCCAGAAATAGAGACGAGTGTGCTGTGTCCGAGCGGCGGAGCTCATCTG 574  
DB 1280 CCTGTATGAGAAACTTGATGAGTGTGAGTGTGCGGAGCGAGGAGCTGTCTG 1339  
QY 575 CTGTACGGCTGCCCTTCCGAGCTTCCACCTGCGCTGTCTCCCTCC 622  
DB 1340 TTGCGCACTTGTTCAGAGTCTTCCATGAGGAGCTGCCATCCACC 1387

RESULT 79  
AAH16238  
ID AAH16238 standard; cDNA; 3079 BP.

AC AAH16238;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15067.  
XX  
KW Human; primer; detection; diagnosis; antilease therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
XX  
PR 27-AUG-1999; 99JP-00300253.  
XX  
PR 11-JAN-2000; 2000JP-00118776.  
XX  
PR 02-MAY-2000; 2000JP-00183767.  
XX  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
XX  
PT Primer sets for synthesising polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNA.

PS Claim 8; SEQ ID NO 15067; 2537BP + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesising 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antilease therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX  
SQ Sequence 3079 BP; 837 A; 744 C; 771 G; 727 T; 0 U; 0 Other;

Query Match 3.2%; Score 48.8; DB 4; Length 3079;  
Best Local Similarity 65.7%; Pred. No. 0.97; Mismatches 37; Indels 0; Gaps 0;  
Matches 71; Conservative 0; Indels 0;

QY 515 CCTGTGCCAGAAATAGAGACGAGTGTGCTGTGTCCGAGCGGCGGAGCTCATCTG 574  
DB 679 CCTGTATGAGAAACTTGATGAGTGTGAGTGTGCGGAGCGAGGAGCTGTCTG 738  
QY 575 CTGTACGGCTGCCCTTCCGAGCTTCCACCTGCGCTGTCTCCCTCC 622  
DB 739 TTGCGCACTTGTTCAGAGTCTTCCATGAGGAGCTGCCATCCACC 786

RESULT 80  
AAK52998  
ID AAK52998 standard; cDNA; 2171 BP.

AC AAK52998;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 2527.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US004098.  
XX  
PR 03-FEB-2000; 2000US-00496914.  
XX  
PR 27-APR-2000; 2000US-00560875.  
XX  
PR 20-JUN-2000; 2000US-00598075.  
XX  
PR 19-JUL-2000; 2000US-00620325.  
XX  
PR 01-SEP-2000; 2000US-00654936.  
XX  
PR 15-SEP-2000; 2000US-00663561.  
XX  
PR 20-OCT-2000; 2000US-00693325.  
XX  
PR 30-NOV-2000; 2000US-00728422.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
XX  
PI Ma Y, Zhao Qh, Wang D, Zhang J, Ren F, Chen R, Wang ZW;  
XX  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;



DR WPI: 2001-476283/51.  
DR P-PSDB; AAM79865.  
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX  
XX  
PS Claim 1; Page 4806-4807; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
SQ Sequence 2171 BP; 349 A; 759 C; 688 G; 375 T; 0 U; 0 Other;  
Query Match 3.1%; Score 48.6; DB 4; Length 2171;  
Best Local Similarity 52.2%; Pred. No. 1;  
Matches 108; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 543 GCCGTGTGCGGAGACGGCGGAGCTCATCTGTGACGCTGCGCTGGGCTTCCAC 602  
DB 91 GCCGTGTGCGGAGCTCATCTGTGACGCTGCGGAGCTCATCTGCTCATCGAT 150  
QY 603 CTGGCCCTGCTGTCCTCCCTCCGCTCCGCGAGATCCCAAGTGAGACCTGGAGTCTCCAGC 662  
DB 151 GCCGCCCGCAAGCTCAAGCGCAAGCGCAAGTCTCCGAGGCGCGCTGCGCCCGCCCGC 210  
QY 663 TGCCTGCAAGCAACAGTCCAGAGAGTGACGCCCGCGGCAAGAGCCCGCCCAAGAG 722  
DB 211 GCCCTTAAGCAACCGCGCAACAGAGACTGGCGCGGCGGCAAGGCGCCCAAGCTG 270  
QY 723 CCACCCGTGAGAGACCCCGCTCCCGCG 749  
DB 271 CCGCCCCCGCAGGCGCCAGCCCGCCGCG 297  
RESULT 81  
AAK52014  
ID AAK52014 standard; cDNA; 2179 BP.  
AC AAK52014;  
XX  
XX  
DT 06-NOV-2001 (first entry)  
DE Human polynucleotide SEQ ID NO 559.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 05-FEB-2001; 2001WO-US004098.  
XX  
XX 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich K;  
DR WPI: 2001-476283/51.  
DR P-PSDB; AAM78881.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX  
XX  
XX Claim 1; Page 2022-2024; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
SQ Sequence 2179 BP; 358 A; 757 C; 689 G; 375 T; 0 U; 0 Other;  
Query Match 3.1%; Score 48.6; DB 4; Length 2179;  
Best Local Similarity 52.2%; Pred. No. 1;  
Matches 108; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 543 GCCGTGTGCGGAGACGGCGGAGCTCATCTGTGACGCTGCGCTGGGCTTCCAC 602  
DB 91 GCCGTGTGCGGAGCTCATCTGTGACGCTGCGGAGCTCATCTGCTCATCGAT 150  
QY 603 CTGGCCCTGCTGTCCTCCCTCCGCTCCGCGAGATCCCAAGTGAGACCTGGAGTCTCCAGC 662  
DB 151 GCCGCCCGCAAGCTCAAGCGCAAGCGCAAGTCTCCGAGGCGCGCTGCGCCCGCCCGC 210  
QY 663 TGCCTGCAAGCAACAGTCCAGAGAGTGACGCCCGCGGCAAGAGACCCCGCCCAAGAG 722  
DB 211 GCCCTTAAGCAACCGCGCAACAGAGACTGGCGCGGCGGCAAGGCGCCCAAGCTG 270  
QY 723 CCACCCGTGAGAGACCCCGCTCCCGCG 749  
DB 271 CCGCCCCCGCAGGCGCCAGCCCGCCGCG 297  
RESULT 82  
AD035478/C  
ID AD035478 standard; DNA; 3005 BP.  
AC AD035478;  
XX  
XX  
DT 26-AUG-2004 (first entry)  
DE Novel mouse gene sequence #151.  
XX  
XX mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;  
KW ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;  
KW viral disorder; ds; gene.  
XX  
XX Mus sp.  
OS  
XX  
XX WO2004046310-A2.  
PN

PD 03-JUN-2004.  
 XX 24-OCT-2003; 2003WO-US033948.  
 XX 15-NOV-2002; 2002US-0426916P.  
 PR 04-DEC-2002; 2002US-0431158P.  
 PR 05-DEC-2002; 2002US-043145P.  
 PR 05-DEC-2002; 2002US-043160P.  
 PR 09-JUN-2003; 2003US-0476621P.  
 PR 09-JUN-2003; 2003US-0476632P.  
 PR 08-JUL-2003; 2003US-0485217P.  
 PR 08-JUL-2003; 2003US-0485359P.  
 PR 08-AUG-2003; 2003US-0493332P.  
 PR 08-AUG-2003; 2003US-0493356P.  
 XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
 XX Williams LT, Chu K, Lee E, Hestir K, Hayashizaki Y, Kamiya M;  
 XX WPI; 2004-431966/40.  
 XX WPI; 2004-431966/40.  
 XX New mouse nucleic acid molecules and polypeptides, useful for treating  
 PT cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart  
 PT disease or thrombosis.  
 XX  
 PS Claim 1, SEQ ID NO 151; 263pp; English.  
 CC The invention comprises 744 novel mouse DNA sequences (genes). The DNA  
 CC sequences of the invention are useful for treating cancer, psoriasis,  
 CC ulcerative colitis, inflammation, ischemic heart disease, thrombosis,  
 CC immune disorders, bacterial disorders and viral disorders. The present  
 CC nucleic acid represents a mouse DNA sequence of the invention. NOTE: The  
 CC present DNA sequence is not shown in the specification, but has been  
 CC retrieved from the WIPO website.  
 SQ Sequence 3005 BP; 782 A; 632 C; 849 G; 742 T; 0 U; 0 Other;  
 Query Match 3.1%; Score 48; DB 12; Length 3005;  
 Best Local Similarity 48.2%; Pred. No. 1.4;  
 Matches 135; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
 QY 1010 GGCCTCCACTTCCGACCGGACCTCCCGGAGCGGCTGCGTGCAGATCTG 1069  
 DB 467 GAGCGCCATGTCGTGGTGGCCGCCACCCCTGCTGCGGCTGCGGCGGCTGCCCGCGG 408  
 QY 1070 CTCAGAGACGTGACCCCAAGCCCTTGTGAGGGGCTGCTGAGCCCGCCCGCCCT 1129  
 DB 407 GCTGGGTGACGAAGCTCTCCGCTGCGGCGGCGGAGCGCTCTAAGCCGAGCCCGGCA 348  
 QY 1130 GGCCTCCGAGCTGCGCAAGATGACACTGCTGACAGAGCCGCTCTGACAGGAGATG 1189  
 DB 347 GGTGATCTCTGCTTGTGGTGGCCGCCGACCGCGCGCTTCAAGCCGAGCGGCGGA 288  
 QY 1190 CCGAGAGTCCCTTCTGAGGAGACACTTTCATGATGATCTCTGATGAGGCGCATCCAG 1249  
 DB 287 GGAAGAGCTTCGCTTACCGACATGCTGAGGCGCGAGCGCGCGCTGTCGGGAG 228  
 QY 1250 CATGCGCGCTCGGCGCGCCCTTCCCTCTGACCCCG 1289  
 DB 227 CCTCGCGCTCTGAGAGCTCGCGGTGCTCCCTCAGCCAG 188  
 RESULT 83  
 AAS75442  
 ID AAS75442 standard; cDNA; 924 BP.  
 XX  
 AC AAS75442;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #11246.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR P-PsDB; ABG11255.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1, SEQ ID NO 11246; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probe, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. Aa64197-Aa59454 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 SQ Sequence 924 BP; 259 A; 192 C; 264 G; 209 T; 0 U; 0 Other;  
 Query Match 3.1%; Score 47.8; DB 5; Length 924;  
 Best Local Similarity 67.7%; Pred. No. 1.3;  
 Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 QY 524 GAAGAATGAGACAGTGTCCGTGTGCGAGCGCGGAGACTATCTGCTGTACGG 583.  
 DB 51 GAGAACTTGATGATGTGAGGTGCGGAGACGAGGAGAGCTGTCTGTTCGACAC 110  
 QY 584 CTGCGCTCGGCGCTTCCACCTGAGCCCTGTCCTCC 622  
 DB 111 TTGTTCAAGATCTTTCATGAGAGTCCACATCCACC 149  
 RESULT 84  
 AAH33830  
 ID AAH33830 standard; cDNA; 1006 BP.  
 XX  
 AC AAH33830;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:886.

```

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US026524.
XX
XX 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX P-PSDB; AAG74399.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 1, Page 2812; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO.1027
XX to 1052, 7921 and 7922
XX
XX Sequence 1006 BP; 339 A; 191 C; 232 G; 239 T; 0 U; 5 Other;
SQ
XX
XX Query Match 3.1%; Score 47.8; DB 4; Length 1006;
XX Best Local Similarity 60.3%; Pred. No. 1.4;
XX Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
XX 528 AATGAGACGAGTGTCCGTGTGTGCGGACGCGGAGCTATCTGCTGTGACGGCTGC 587
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 233 AATGAAGACTGTGTGCTGTGCTGCCAAAACGGAGGAGATCTTGTGCTGCCGAAAATGT 292
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 588 CCTCGGGCCCTTCACCTGGGCTGCTGTGCTCCGCTCGCGGAGATCCCACTGGGAC 647
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 293 CCAAGAGCTTTTCATCTTAACCTGTCTCAATGTTCCAACTTAAGCTTTCCAACTGGGAC 352
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 648 TGGAGGTGCTC 658
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 353 TGGATATGCAC 363
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 85
XX AAH16455
XX ID AAH16455 standard; cDNA; 2724 BP.
XX
XX AC AAH16455;
XX
XX DT 26-JUN-2001 (first entry)
XX

```

```

DB Human cDNA sequence SEQ ID NO:15457.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 15457; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 2724 BP; 905 A; 602 C; 590 G; 627 T; 0 U; 0 Other;
SQ
XX
XX Query Match 3.1%; Score 47.8; DB 4; Length 2724;
XX Best Local Similarity 60.3%; Pred. No. 1.6;
XX Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
XX 528 AATGAGACGAGTGTCCGTGTGTGCGGACGCGGAGCTATCTGCTGTGACGGCTGC 587
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1953 AATGAAGACTGTGTGCTGTGCTGCCAAAACGGAGGAGATCTTGTGCTGCCGAAAATGT 2012
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 588 CCTCGGGCCCTTCACCTGGGCTGCTGTGCTCCGCTCGCGGAGATCCCACTGGGAC 647
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2013 CCAAGAGCTTTTCATCTTAACCTGTCTCAATGTTCCAACTTAAGCTTTCCAACTGGGAC 2072
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 648 TGGAGGTGCTC 658
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2073 TGGATATGCAC 2083
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX

```



Db 2800 CCAAGGCTTTCATCTAATCTGTATCTCCACACTACTTACGTTTCCAGTGGGAC 2859  
 QY 648 TGGAGTGCTC 658  
 Db 2860 TGGATATGCAC 2870

## RESULT 88

ADL22564  
 ID ADL22564 standard; cDNA; 6730 BP.

AC ADL22564;

XX 20-MAY-2004 (first entry)

DE Human disease detection and treatment (MDDT) cDNA - SEQ ID 13.

XX disease detection; MDDT; antiarteriosclerotic; antipsoriatic; cytostatic;

KW cell signalling; arteriosclerosis; psoriasis; cancer; human; ss; gene.

XX Homo sapiens.

XX WO2003062379-A2.

PD 31-JUL-2003.

XX 14-JAN-2003; 2003WO-US001363.

XX 17-JAN-2002; 2002US-0349413P.

XX 17-JAN-2002; 2002US-0349946P.

XX (INCY-) INCYTE GENOMICS INC.

PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL,

PI Yu JY, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TF, Gerstin EH,

PI Peralta CH, Lewis SA, Chen AJ, Marwaha R, Lan RX, Urashka ME,

PI Kristenam SR, Koliuru V, Panesar IS;

XX MPI; 2003-853443/79.

DR P-PSDB; ADL22668.

PT New isolated disease detection and treatment polynucleotide for  
 PT diagnosing or treating conditions associated with cell signaling e.g.  
 PT arteriosclerosis, psoriasis, and cancer.

PS Claim 1; SEQ ID NO 13; 411pp; English.

XX The invention relates to a novel isolated disease detection and treatment  
 CC (MDMT) DNA polynucleotide. The polynucleotide of the invention  
 CC demonstrates antiarteriosclerotic, antipsoriatic and cytostatic  
 CC activities and may be useful in a composition for detecting the  
 CC expression of a disease detection and treatment molecule polynucleotide.  
 CC The molecules of the invention may be utilised to diagnose or treat  
 CC conditions, diseases or disorders associated with cell signalling, such  
 CC as arteriosclerosis, psoriasis and cancer. The current sequence is that  
 CC of a human MDDT cDNA of the invention.

XX Sequence 6730 BP; 2123 A; 1184 C; 1206 G; 2217 T; 0 U; 0 Other;

Query Match 3.1%; Score 47.8; DB 11; Length 6730;

Best Local Similarity 60.3%; Pred. No. 1.8;

Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 528 AATGAGGAGGAGTGTCCTGTGTGGAGCGCGGGAGCTCATCTGTGTGACGGCTGC 587

Db 1078 AATGAAGACTGTGTCTGTCTGTCCAAAAGGAGGAGATCTTGTGTGCGGAAAATGT 1137

QY 588 CCGCGGGCTTCCACTGGCGCTGTCCCTCCGCTCCGGAGANCCCAAGTGGAGC 647

Db 1138 CCAAGGCTTTCATCTAATCTGTATCTCCACACTACTTACGTTTCCAGTGGGAC 1197

QY 648 TGGAGTGCTC 658

Db 1198 TGGATATGCAC 1208

## RESULT 89

ABL06755  
 ID ABL06755 standard; cDNA; 3223 BP.

XX ABL06755;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14747.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX MPI; 2001-656860/75.

DR P-PSDB; ABB62652.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Claim 1; SEQ ID NO 14747; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3223 BP; 793 A; 935 C; 890 G; 605 T; 0 U; 0 Other;

Query Match 3.1%; Score 47.6; DB 4; Length 3223;

Best Local Similarity 55.9%; Pred. No. 1.8;

Matches 114; Conservative 0; Mismatches 84; Indels 6; Gaps 1;

QY 526 AGAATGAGGAGGAGTGTCCTGTGTGGAGCGCGGGAGCTCATCTGTGTGACGCT 585

Db 202 ACAACCAACACTATTTGTGACGCTCGAGAGGGGGCAATGTGCTGTGCGATCGAT 261

QY 586 GCCCTCGGGCTTCCACTGGCGCTGTGCCCTCCGCTCCGGAG-----ATCCCA 639

Db 262 GCCCTCCAGCTTTCACCTGCAATGCAATGATCCACCTTTGAGGAGAGGACATACCA 321

QY 640 GTGGGACCTGAGGTGCTTCCAGCTGCTGACGCAACATTCAGAGAGTGACGCCGGG 699

Db 322 GTGGGCAAGTGGGTGGGACCAAGCTGTGCGATGAGAGAGCTTCCAGCGCGGCTCT 381

QY 700 CAGAGAGCGCGGCGCCAGAGGC 723

Db 382 CATCCAGGCGCAGTTCCGTGAGC 405

```

RESULT 90
AAV24559
ID AAV24559 standard; DNA; 2905 BP.
XX
XX AAV24559;
AC
XX 17-SEP-1998 (first entry)
XX
XX Leukocyte specific protein, Sp140, coding sequence.
DE Leukocyte specific protein, Sp140, coding sequence.
XX
XX Sp140; leukocyte specific protein; gene transcription regulator; therapy;
XX autoimmune disease; viral infection; cancer; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 107..2368
XX /tag a
XX /product= "Sp140"
XX
XX WO9814569-A1.
XX
XX 09-APR-1998.
XX
XX 02-OCT-1997; 97WO-US017715.
XX
XX 02-OCT-1996; 96US-0027347P.
XX
XX (BLOC/) BLOC D B.
XX (BLOC/) BLOC K D.
XX
XX Bloch DB, Bloch KD;
XX
XX MPI: 1998-286419/25.
XX P-PSDB; AAW57747.
XX
XX New isolated gene transcription regulator, Sp140 - used to develop
XX products for the diagnosis and treatment of auto-immune diseases, viral
XX infections or cancers.
XX
XX Claim 1, Page 54-57; 81pp; English.
XX
XX This sequence encodes the leukocyte specific protein, Sp140 of the
XX invention. The Sp140 polypeptides act as gene transcription regulators.
XX They can be used to develop products for use in the diagnosis and
XX treatment of autoimmune diseases such as primary biliary cirrhosis,
XX rheumatoid arthritis, systemic lupus erythematosus, Sjogren's syndrome,
XX scleroderma and multiple sclerosis, viral diseases including those caused
XX by herpes simplex virus, cytomegalovirus, HIV, hepatitis virus, human T-
XX cell leukemia virus-1 (HTLV-1) and adenovirus, and cancers including
XX leukaemia, particularly acute promyelocytic leukaemia, cancers of the
XX breast, ovary, prostate, bone, liver, pancreas or spleen, sarcomas and
XX melanomas
XX
XX Sequence 2905 BP; 874 A; 610 C; 760 G; 661 T; 0 U; 0 Other;
SQ
XX
XX Query Match 3.1%; Score 47.2; DB 2; Length 2905;
XX Best Local Similarity 64.8%; Pred. No. 2.1;
XX Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 515 CCTGTGCCAGAAAGATGAGACGAGTGTGCGGTGTCCGAGCGCGGAGCTCATCTG 574
DB 1816 CCTGTATGAGAAACCTGATGAGTGTGCGGTGTCCGAGCGGAGCTCATCTG 1875
QY 575 CTGTGACGAGCTGCGCTCGGAGCTTCACCTGCGCTGCTCCCTCC 622
DB 1876 TTGCGACACTTGTCAAGAGCTTCCATGAGACGTCTCACCCTCC 1923

```

RESULT 91  
AA575444

```

ID AAS75444 standard; cDNA; 2911 BP.
XX
XX AAS75444;
AC
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #11248.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSB-) HYSBQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX P-PSDB; ABG11257.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 11248; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostic as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2911 BP; 878 A; 609 C; 759 G; 665 T; 0 U; 0 Other;
SQ
XX
XX Query Match 3.1%; Score 47.2; DB 5; Length 2911;
XX Best Local Similarity 64.8%; Pred. No. 2.1;
XX Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 515 CCTGTGCCAGAAAGATGAGACGAGTGTGCGGTGTCCGAGCGCGGAGCTCATCTG 574
DB 1816 CCTGTATGAGAAACCTGATGAGTGTGCGGTGTCCGAGCGGAGCTCATCTG 1875
QY 575 CTGTGACGAGCTGCGCTCGGAGCTTCACCTGCGCTGCTCCCTCC 622
DB 1876 TTGCGACACTTGTCAAGAGCTTCCATGAGACGTCTCACCCTCC 1923

```

## RESULT 92

AA575443  
ID AA575443 standard; cDNA; 3252 BP.

AC AA575443;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11247.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

PI MPI; 2001-639362/73.

DR P-PSDB; ABG11256.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PT Claim 1; SEQ ID NO 11247; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (II) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 3252 BP; 987 A; 678 C; 866 G; 721 T; 0 U; 0 Other;

SQ Query Match 3.1%; Score 47.2; DB 5; Length 3252;

Best Local Similarity 64.8%; Pred. No. 2.1;

Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 515 CCTGTCCCAAGATGATGACGAGTGTGCTGTCCGAGCGCGGAGCTCATCTG 574

DB 2168 CCTTTATATGAGAAACCTGATGATGTGATGATGTGCTCCGAGCGAGGAGCTTTCTG 2227

QY 575 CTGTGACGGAGCTGCTCGGAGCTTTCACCTGAGCTCTGTCCCTCC 622

DB 2228 TTGCGACACTTGTTCAGAGTCTTCATGAGACTGTCAATCCGCC 2275

## RESULT 93

AA575445  
ID AA575445 standard; cDNA; 1210 BP.

AC AA575445;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11249.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

PI MPI; 2001-639362/73.

DR P-PSDB; ABG11258.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PT Claim 1; SEQ ID NO 11249; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (II) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1210 BP; 248 A; 321 C; 368 G; 273 T; 0 U; 0 Other;

SQ Query Match 3.0%; Score 47; DB 5; Length 1210;

Best Local Similarity 57.9%; Pred. No. 2.1;

Matches 106; Conservative 0; Mismatches 70; Indels 7; Gaps 1;

QY 524 GAGAGATGAGAGCGAGTGTGCTGTGTCGAGCGCGGAGCTCATCTGCTGAGCG 583

DB 588 GAGAAACCTGATGATGATGTGAGCTGTGCGGAGCGAGAGCTTTTGTGCGACAC 647

QY 584 CTGCCCTGCGGAGCTTTCACCTGAGCTGTGCTGTCCCTCC-----GCTCCGAGAGATCC 636



DB 648 TTGTTCAAGAGCTTTCATGAGAGCTGTACATCCCGCTGTGGAAGCTGAGAGGATCT 707  
QY 637 CCAATGGAGACCTGAGAGTGTCTCAGCTGCTGCAGGCAACAGTCCAGAGAGTGACAGCC 696  
DB 708 TCACACTGTCATGCGCAGCTCTCCAGCAGAGTGCCGCTGTGTCTGATGGGAAAGAC 767  
QY 697 GGG 699  
DB 768 AGG 770

RESULT 94  
AAS29014  
ID AAS29014 standard; DNA; 6610 BP.  
XX  
AC AAS29014;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Genomic sequence #20 encoding novel human uterine motility polypeptide.  
XX  
KW Human; uterine motility-association disorder; uterus; pregnancy; labour;  
XX menstrual cycle; gene therapy; ds.  
OS Homo sapiens.  
XX  
PN WO200155201-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001317.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-021647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-022547P.  
PR 14-AUG-2000; 2000US-022557P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-022679P.  
PR 22-AUG-2000; 2000US-022681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241251P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0244755P.  
PR 08-NOV-2000; 2000US-0244756P.  
PR 08-NOV-2000; 2000US-024477P.  
PR 08-NOV-2000; 2000US-024478P.  
PR 08-NOV-2000; 2000US-0245223P.  
PR 08-NOV-2000; 2000US-0245224P.  
PR 08-NOV-2000; 2000US-024524P.  
PR 08-NOV-2000; 2000US-0245255P.  
PR 08-NOV-2000; 2000US-0245256P.  
PR 08-NOV-2000; 2000US-024527P.  
PR 08-NOV-2000; 2000US-024528P.  
PR 08-NOV-2000; 2000US-024532P.  
PR 08-NOV-2000; 2000US-024609P.  
PR 08-NOV-2000; 2000US-024610P.  
PR 08-NOV-2000; 2000US-024611P.  
PR 08-NOV-2000; 2000US-024613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.



```

PR 08-SEP-2000; 2000US-0232081P
PR 12-SEP-2000; 2000US-0231968P
PR 14-SEP-2000; 2000US-0232397P
PR 14-SEP-2000; 2000US-0232398P
PR 14-SEP-2000; 2000US-0232399P
PR 14-SEP-2000; 2000US-0232400P
PR 14-SEP-2000; 2000US-0232401P
PR 14-SEP-2000; 2000US-0233063P
PR 14-SEP-2000; 2000US-0233064P
PR 21-SEP-2000; 2000US-0233655P
PR 21-SEP-2000; 2000US-0234223P
PR 21-SEP-2000; 2000US-0234274P
PR 25-SEP-2000; 2000US-0234977P
PR 25-SEP-2000; 2000US-0234988P
PR 26-SEP-2000; 2000US-0235484P
PR 27-SEP-2000; 2000US-0235834P
PR 29-SEP-2000; 2000US-0236327P
PR 29-SEP-2000; 2000US-0236367P
PR 29-SEP-2000; 2000US-0236368P
PR 29-SEP-2000; 2000US-0236369P
PR 29-SEP-2000; 2000US-0236370P
PR 02-OCT-2000; 2000US-0236802P
PR 02-OCT-2000; 2000US-0237037P
PR 02-OCT-2000; 2000US-0237038P
PR 02-OCT-2000; 2000US-0237039P
PR 13-OCT-2000; 2000US-0237040P
PR 13-OCT-2000; 2000US-0239335P
PR 20-OCT-2000; 2000US-0239337P
PR 20-OCT-2000; 2000US-0240960P
PR 20-OCT-2000; 2000US-0241221P
PR 20-OCT-2000; 2000US-0241785P
PR 20-OCT-2000; 2000US-0241786P
PR 20-OCT-2000; 2000US-0241787P
PR 20-OCT-2000; 2000US-0241808P
PR 20-OCT-2000; 2000US-0241809P
PR 01-NOV-2000; 2000US-0244617P
PR 08-NOV-2000; 2000US-0246474P
PR 08-NOV-2000; 2000US-0246475P
PR 08-NOV-2000; 2000US-0246476P
PR 08-NOV-2000; 2000US-0246477P
PR 08-NOV-2000; 2000US-0246478P
PR 08-NOV-2000; 2000US-0246523P
PR 08-NOV-2000; 2000US-0246524P
PR 08-NOV-2000; 2000US-0246525P
PR 08-NOV-2000; 2000US-0246526P
PR 08-NOV-2000; 2000US-0246527P
PR 08-NOV-2000; 2000US-0246528P
PR 08-NOV-2000; 2000US-0246532P
PR 08-NOV-2000; 2000US-0246609P
PR 08-NOV-2000; 2000US-0246610P
PR 08-NOV-2000; 2000US-0246611P
PR 08-NOV-2000; 2000US-0246613P
PR 17-NOV-2000; 2000US-0249207P
PR 17-NOV-2000; 2000US-0249208P
PR 17-NOV-2000; 2000US-0249209P
PR 17-NOV-2000; 2000US-0249210P
PR 17-NOV-2000; 2000US-0249211P
PR 17-NOV-2000; 2000US-0249212P
PR 17-NOV-2000; 2000US-0249213P
PR 17-NOV-2000; 2000US-0249214P
PR 17-NOV-2000; 2000US-0249215P
PR 17-NOV-2000; 2000US-0249216P
PR 17-NOV-2000; 2000US-0249217P
PR 17-NOV-2000; 2000US-0249218P
PR 17-NOV-2000; 2000US-0249244P
PR 17-NOV-2000; 2000US-0249245P
PR 17-NOV-2000; 2000US-0249246P
PR 17-NOV-2000; 2000US-0249247P
PR 17-NOV-2000; 2000US-0249248P
PR 17-NOV-2000; 2000US-0249249P
PR 17-NOV-2000; 2000US-0249300P

```

```

PR 01-DEC-2000; 2000US-0250160P
PR 01-DEC-2000; 2000US-0250391P
PR 05-DEC-2000; 2000US-0251030P
PR 05-DEC-2000; 2000US-0251988P
PR 05-DEC-2000; 2000US-0256719P
PR 08-DEC-2000; 2000US-0251479P
PR 08-DEC-2000; 2000US-0251856P
PR 08-DEC-2000; 2000US-0251868P
PR 08-DEC-2000; 2000US-0251869P
PR 08-DEC-2000; 2000US-0251989P
PR 11-DEC-2000; 2000US-0254097P
PR 05-JAN-2001; 2001US-0259678P
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SW,
XX WPI; 2001-488777/53.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Disclosure; SEQ ID NO 147; 524bp; English.
XX
XX The present invention relates to the isolation of novel human uterine
XX motility-associated polypeptides (AAU18094-AAU18152), and cDNA and
XX genomic sequences encoding for these polypeptides. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with uterine motility such as pregnancy
XX and labour, and menstrual disorders. The polynucleotide sequences of the
XX invention are also useful in gene therapy. AAS28995-AAS29020 represent
XX genomic sequences encoding for novel human uterine motility-associated
XX polypeptides. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPo at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;
SQ
Query Match 3.0%; Score 47; DB 4; Length 6610;
Best Local Similarity 53.6%; Pred. No. 2.6;
Matches 143; Conservative 0; Mismatches 120; Indels 4; Gaps 2;
QY 946 GCGGCGTGTGCGAGATGTACGACGTGCGGTACTGACGTGCGCGCTGCTTC 1005
DB 7 GCTGTACGTGCGCGAGCGTACGTTGCGGCGCGCTGCGATGCTACGTCAGTCTCC 66
QY 1006 ACTGGCGCTGCACCTTCCAGCGCGGCACTCCCGGCGCGGCGCTGCGTCAAGT 1065
DB 67 GCGGCGCTGCACCGCTGTGACGCGCAGCTGTGCTGCGAGCGCGGCGTCAAGT 126
QY 1066 CTTGCTCAGGAGACGTGACCGCCCGCTGTGAGGCGGTGCTGAGCCCGCAGCCGCC 1125
DB 127 ACTG-TGCGGAGCGCTGCGCGCGCGCTTCTACGCTTGCGTGTGCGG---CCGATA 182
QY 1126 GCGTGGCCCTTGCGGCTGTGCAAGATGACACTGCGAGTACGAGCGCGCTTGCACAGG 1185
DB 183 GCGGCGTCTCTGCGCAGGAGGAGAGGTGGAAGGCGGCGCGCGATGAGGACAGGG 242
QY 1186 ATGACCTGAGTCTCTTCTGAGCGAGC 1212
DB 243 GTGAGCGCGGTCCAGAGCTTCCCTTACG 269
RESULT 96
AAS26967
ID AAS26967 standard; DNA; 6610 BP.
XX AAS26967;
AC AAS26967;
XX
XX 07-NOV-2001 (first entry)
XX

```

DE Human genomic DNA encoding partial novel secreted protein, Seq ID 303.  
XX  
KW Human; immunosuppressive; antiarthritic; ds; antineumatic; cytostatic;  
KW caritant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; opthalmological; vlnetary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disease; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO200155441-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001320.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228927P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234972P.  
PR 25-SEP-2000; 2000US-0234978P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241825P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249247P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.

PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251889P.  
PR 08-DEC-2000; 2000US-0251909P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-476222/51.  
XX  
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to  
PT diagnose diseases or disorders associated with aberrant expression or  
PT activity of polypeptides, for treating blood clotting disorder,  
PT hemophilia.  
XX  
XX Disclosure; SEQ ID NO 303; 601bp; English.

XX The invention relates to isolated nucleic acid molecules and their  
XX encoded secreted proteins. The nucleic acids and proteins are used to  
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
XX in diagnosing a pathological condition or susceptibility to a  
XX pathological condition. Antibodies to the proteins can also be used in  
XX alleviating symptoms associated with the disorders and in diagnostic  
XX immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
XX (ELISA). Disorders which are diagnosed or treated include autoimmune  
XX diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
XX neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
XX arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
XX nervous system disorders e.g. Alzheimer's disease, infections caused by  
XX bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
XX and many other disorders listed in the specification. The polypeptides  
XX can also be used to aid wound healing and epithelial cell proliferation,  
XX to prevent skin aging due to sunburn, to maintain organs before  
XX transplantation, for supporting cell culture of primary tissues, to  
XX regenerate tissues and in chemotaxis. The polypeptides can also be used  
XX as a food additive or preservative to increase or decrease storage  
XX capabilities, fat content, lipid, protein, carbohydrate, vitamin,  
XX minerals, cofactors and other nutritional components. The present  
XX sequence is a genomic DNA encoding a partial novel secreted protein of  
XX the invention. Note: The sequence data for this patent did not form part  
XX

Query Match 3.0%; Score 47; DB 4; Length 6610;

Best Local Similarity 53.6%; Pred. No. 2.6;

Matches 143; Conservative 0; Mismatches 120; Indels 4; Gaps 2;

OY 946 GCGGGGTGGGAGATGAGTGAAGAGTGTCTGTGGGTTACTACATGCGCGCTTC 1005  
DB 7 GCTGTACGTGCGCGAGGCTTCTGCGCGCGCTGCGATCTACATGCGCATGCTTC 66  
OY 1006 ACTGGGCTGACCACTTCCAGCGGACCTCCCGGCGGAGAGCGGCTGCGTGCAGAT 1065  
DB 67 GCGGCGGCTGCCACCTGTGAGACGCACTGTGCGGCTGCGGCTACCGGCGCAAT 126  
OY 1066 CTGTCTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1125  
DB 127 ACTG-TGCGGAGACCGGCGCGCGCTTCTACGCGCTTGTGGCTGTGCGG--CCGATA 182  
OY 1126 GCTGGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1185  
DB 183 GCGGCGGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242  
OY 1186 ATGACCTGAGTCCCTTCTGAGGAGC 1212  
DB 243 GTGAGGCGGCTCCAGACCTTCCCTAGC 269

RESULT 97  
AAS26968  
ID AAS26968 standard; DNA; 6610 BP.  
XX  
XX AAS26968;  
AC  
XX 07-NOV-2001 (first entry)  
DT  
XX  
XX Human genomic DNA encoding partial novel secreted protein, Seq ID 304.  
DE  
XX  
XX Human; immunosuppressive; antiarthritic; ds; antineuritic; cytosolic;  
XX Human; vasotropic; cerebroprotective; neurotropic; neuroprotective;  
XX antibacterial; virucide; fungicide; opthalmological; vulnary;  
XX secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
XX cerebral ischaemia; angiogenesis; nervous system disorder;  
XX Alzheimer's disease; infection; ocular disorder; corneal infection;  
XX wound healing; epithelial cell proliferation; skin ageing; food additive;  
XX preservative; antiproliferative.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200155441-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001320.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216447P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0216290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228927P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.



OY 1126 GCCTGGCCCCCTGGGCTGCCAGATGACACTGCCAGTACGAGCCCCCTTGACACAGG 1185  
DB 183 GGGCGGCTCTCTCGGAGGAGAGAGAGGTGAGAGGGGGGGCCCGGATGGGGACAGG 242  
OY 1186 ATGACCTGGAGTCCCTTCTGAGGAGC 1212  
DB 243 GTGAGGGGGGTCCACGCTTCCCTTACG 269

RESULT 98  
ABA06837  
ID ABA06837 strand: DNA; 6610 BP.  
XX ABA06837;  
XX  
XX  
DT 10-JAN-2002 (first entry)  
XX  
DE Human genomic DNA SEQ ID NO: 923.  
XX  
XX Human; gene therapy; neural disorder; immune system disorder;  
KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
KW pulmonary disorder; cardiovascular disorder; renal disorder;  
KW proliferative disorder; inflammation; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200154474-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001349.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225577P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 23-AUG-2000; 2000US-0227082P.  
PR 30-AUG-2000; 2000US-0228824P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239336P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244674P.  
PR 08-NOV-2000; 2000US-0244755P.  
PR 08-NOV-2000; 2000US-0244756P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246603P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.



```
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0253678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476161/51.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
XX peptide is used in preventing, treating or ameliorating a medical
XX condition.
XX
XX PS Disclosure; SEQ ID NO 923; 859pp + Sequence Listing; English.
XX
XX CC The present invention provides human cDNAs, proteins and related genomic
XX cDNAs. These can be used in the treatment of neural, immune system,
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
XX renal and proliferative disorders and inflammation. The present sequence
XX is a genomic DNA of the invention
XX
XX SQ Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;
XX
XX Query Match 3 0%; Score 47; DB 4; Length 6610;
XX Best Local Similarity 53.6%; Pred. No. 2.6; Mismatches 120; Indels 4; Gaps 2;
XX Matches 143; Conservative 0;
QY 946 GCGGGGTTGCGAGATGTAGTACGACGTGCTGCGGTGTACTGACGCGCGCTTCC 1005
DB 7 GCTGTACAGTCCGCGAGCGTACGTTGCGGCGCGCTGCGATGCTACTGCTCAGTGTCTCC 66
QY 1006 ACTGGGCTGTCCACTTCCAGCCGGGACCTCCCGCGCGGAGCGGCGCTTCCGCTGACAT 1065
DB 67 GGGGGCGCTGCACCCCTGTGACGCGACGTGTGCTTGCAGCGCGGCTTACCGCGCAAGT 126
QY 1066 CCTGCTCAGAGAGAGTACCCAGCCCGCTGTGAGAGGGGTGTGGCCCCCGACCGCGCC 1125
DB 127 ACTG-TTCGGAGACCGTGGCCCGCGGCTTCTACGGCTTGGGCTGTGCGCG---CCGGTAA 182
QY 1126 GCGTGGCCCTTGGCGCTGCGCAAGATGACATGTACATGACAGCGCCGCTTGCACAGGG 1185
DB 183 GCGCGGTCTCTCGGCGCAGGAGGAGAGTGAAGGAGGCGCGCGCGCGATGGGACAGGG 242
QY 1186 ATGACCTGAGATTCCTTCTGAGCGAGC 1212
DB 243 GTGAGCGGGTCCAGCGCTTCCCTTACG 269
XX
XX RESULT 99
XX ABA06838 standard; DNA; 6610 BP.
XX
```

```
AC ABA06838;
XX
XX 10-JAN-2002 (first entry)
XX
XX DE Human genomic DNA SEQ ID NO: 924.
XX
XX KW Human; gene therapy; neural disorder; immune system disorder;
XX muscular disorder; reproductive disorder; gastrointestinal disorder;
XX pulmonary disorder; cardiovascular disorder; renal disorder;
XX proliferative disorder; inflammation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200154474-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US001349.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225475P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
XX PR 22-AUG-2000; 2000US-0226681P.
XX PR 22-AUG-2000; 2000US-0226686P.
XX PR 22-AUG-2000; 2000US-0226868P.
XX PR 22-AUG-2000; 2000US-0227182P.
XX PR 23-AUG-2000; 2000US-0227109P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
XX PR 05-SEP-2000; 2000US-0229509P.
XX PR 05-SEP-2000; 2000US-0229513P.
XX PR 06-SEP-2000; 2000US-0230437P.
XX PR 06-SEP-2000; 2000US-0230438P.
XX PR 06-SEP-2000; 2000US-0230439P.
XX PR 08-SEP-2000; 2000US-0231242P.
XX PR 08-SEP-2000; 2000US-0231243P.
XX PR 08-SEP-2000; 2000US-0231244P.
XX PR 08-SEP-2000; 2000US-0231413P.
XX PR 08-SEP-2000; 2000US-0231414P.
XX PR 08-SEP-2000; 2000US-0232080P.
XX PR 08-SEP-2000; 2000US-0232081P.
XX PR 12-SEP-2000; 2000US-0231968P.
XX PR 14-SEP-2000; 2000US-0232397P.
XX PR 14-SEP-2000; 2000US-0232398P.
```



KM respiratory; renal; infectious disease; gastrointestinal; gene therapy;  
KM neuronal growth; neuronal disorder; neuro-degenerative condition;  
KM keratinocyte growth; human; ds.  
XX Homo sapiens.  
XX OS  
XX US2002086330-A1.  
XX  
XX 04-JUL-2002.  
XX  
PF 17-JAN-2001; 2001US-00764893.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220863P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 22-AUG-2000; 2000US-0226688P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 05-SEP-2000; 2000US-0229509P.  
XX 08-SEP-2000; 2000US-0229513P.  
XX 08-SEP-2000; 2000US-0231413P.  
XX 21-SEP-2000; 2000US-0234223P.  
XX 21-SEP-2000; 2000US-0234274P.  
XX 25-SEP-2000; 2000US-0234979P.  
XX 27-SEP-2000; 2000US-0235834P.  
XX 29-SEP-2000; 2000US-0236327P.  
XX 29-SEP-2000; 2000US-0236367P.  
XX 29-SEP-2000; 2000US-0236368P.  
XX 29-SEP-2000; 2000US-0236369P.  
XX 29-SEP-2000; 2000US-0236370P.  
XX 02-OCT-2000; 2000US-0236802P.  
XX 02-OCT-2000; 2000US-0237037P.  
XX 02-OCT-2000; 2000US-0237038P.  
XX 02-OCT-2000; 2000US-0237039P.  
XX 02-OCT-2000; 2000US-0237040P.  
XX 13-OCT-2000; 2000US-0239335P.  
XX 20-OCT-2000; 2000US-0240960P.  
XX 20-OCT-2000; 2000US-0241785P.  
XX 20-OCT-2000; 2000US-0241809P.  
XX 01-NOV-2000; 2000US-0244617P.  
XX 17-NOV-2000; 2000US-0249299P.  
XX 08-DEC-2000; 2000US-0251856P.  
XX 08-DEC-2000; 2000US-0251868P.  
XX 08-DEC-2000; 2000US-0251869P.  
XX  
XX (ROSE/) ROSEN C A.  
XX (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX  
XX MPI; 2002-665432/71.  
XX  
XX Novel polypeptide useful for diagnosis, prognosis, prevention, and  
XX treatment of immune, hyperproliferative, renal, respiratory, and  
XX PT

PT cardiovascular, reproductive, endocrine, gastrointestinal and  
PT neurological disorders.  
XX  
XX Disclosure, Page 323-326; 335pp; English.  
XX  
XX The invention relates to an isolated polypeptide comprising a sequence at  
XX least 90% identical to a full length protein sequence selected from 55  
XX CC sequences given in the specification such as a sequence of 163, 74 or 140  
XX CC amino acids fully defined in the specification, or the encoding sequence  
XX CC contained in 49 cDNA clones given in specification e.g. HCFAT05, HMA895  
XX CC or HTNEM01. The protein and its encoding nucleic acid are useful for  
XX CC diagnosing a pathological condition or susceptibility to a pathological  
XX CC condition in a subject and for preventing, treating or ameliorating a  
XX CC medical condition. The protein, its encoding nucleic acid and an isolated  
XX CC antibody that can bind to the protein are useful in treating, preventing,  
XX CC diagnosing and/or prognosing immunodeficiencies, autoimmune disorders,  
XX CC allergic reactions and conditions, inflammatory conditions, graft-versus-  
XX CC host disease, blood-related disorders, hyperproliferative disorders,  
XX CC renal disorders, cardiovascular disorders, respiratory disorders,  
XX CC neurological disorders, endocrine disorders, reproductive system  
XX CC disorders, infectious diseases, and gastrointestinal disorders. The  
XX CC protein of the invention is useful to stimulate neuronal growth and to  
XX CC treat, prevent, and/or diagnose neuronal damage which occurs in certain  
XX CC neuronal disorders or neuro-degenerative conditions, for stimulating  
XX CC keratinocyte growth, to prevent hair loss, to modulate mammalian  
XX CC characteristics such as body height, weight, hair color, and to increase  
XX CC or decrease storage capabilities, fat content, lipid, protein,  
XX CC carbohydrate, vitamins, minerals, cofactors or other nutritional  
XX CC components. The nucleic acid of the invention can be used in gene  
XX CC therapy. This polynucleotide sequence represents one of the novel nucleic  
XX CC acids of the invention  
XX  
SQ Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;  
XX  
Query Match 3.0%; Score 47; DB 6; Length 6610;  
Best Local Similarity 53.6%; Pred. No. 2.6;  
Matches 143; Conservative 0; Mismatches 120; Indels 4; Gaps 2;  
QY 946 GCGGGGTGGGGAATGATGTAACGACGTCGCGGTGTAACGACGCGCGCTGCCCTCC 1005  
DB 7 GCTGTCAAGTCCGCGAGCGTACGTTCGGCGCGCGCTGCGATGCTAATGCCAGTCTTCC 66  
QY 1006 ACTGGCGTGCACCTTCCAGCCGACCTCCCGGCCGCGAGCGGAGCTGCGCTGCAGAT 1065  
DB 67 GCGGCGGCTGCACCTTCCAGCGGACGTCGTCGAGAGCGGAGCTGCGCGGCAAGT 126  
QY 1066 CTTGCTCAGAGAGCTGACCCCAAGCCCTGTGAGAGGGGTGCTGCCCCCAAGCCCCGCC 1125  
DB 127 ACTG-TTCGAGCCCTGCCCCCGCTTCTAAGCTTGGGCTGTGCGCG--CCGATGA 182  
QY 1126 GCGTGGCCCCCTGGGCTGCGCAAGATGACATGCGACGACGAGCCGCTGTCACAGGG 1185  
DB 183 GCGGCGTCTTCGCGCAAGAGAGAGAGGTGAAGAGCGGCGGCCCGATGAGGACAGGG 242  
QY 1186 ATGACCTGAGTCCCTTCTGAGCGAGC 1212  
DB 243 GTGAGCGCGGATCCAGCTTCCCTAGC 269

Search completed: February 9, 2005, 14:33:16  
Job time : 768 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 157 Seconds  
(without alignments)

6994.706 Million cell updates/sec

Title: US-09-508-658A-3

Sequence: 1 agagaagaagcagagctcttc.....aaaataataaataacgctg 1545

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum first 200 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/1/ina/PCTUS91-06418-2
- 6: /cgn2\_6/prodata/1/ina/backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	3.9	6328	3	US-08-913-832A-1
2	60	3.9	6328	4	US-09-249-181A-1
3	60	3.9	6328	4	US-09-158-707-1
4	60	3.9	6475	4	US-09-620-3120-325
5	59.8	3.9	1053	4	US-07-945-295-1
6	59.8	3.9	1053	5	PCT-US91-06418-2
7	47.2	3.1	2905	3	US-08-942-686-1
8	44.6	2.9	2373	4	US-09-976-594-1064
9	44.6	2.9	4616	1	US-08-340-203A-1
10	44.6	2.9	4616	2	US-08-452-567-1
11	44.6	2.9	4616	2	US-08-452-427-1
12	44.6	2.9	4616	3	US-09-085-407-1
13	42.6	2.8	882	3	US-09-920-923B-36
14	42.6	2.8	960	4	US-09-526-400-3
15	42.6	2.8	1182	4	US-09-526-400-1
16	42.6	2.8	1469	5	PCT-US92-08258-1
17	42.2	2.7	3096	4	US-09-270-767-13712
18	42.2	2.7	4403765	3	US-09-103-840A-2
19	42.2	2.7	4411529	3	US-09-103-840A-1
20	42	2.7	3612	4	US-09-387-286-60
21	41.6	2.7	1941	5	PCT-US91-07635-5
22	41.4	2.7	7218	1	US-08-332-463-14
23	41.2	2.7	14707	1	US-09-312-762A-3
24	40.8	2.6	364	3	US-09-103-840A-1
25	40.6	2.6	430	4	US-09-621-976-1702
26	40.4	2.6	3033	4	US-09-621-976-1656
27	40.4	2.6	3033	4	US-09-724-797-81

28	40.2	2.6	1556	2	US-08-881-857-1	Sequence 1, Appl
29	40.2	2.6	1556	3	US-09-233-342A-1	Sequence 1, Appl
30	40.2	2.6	1682	4	US-09-620-312D-389	Sequence 389, App
31	40.2	2.6	2399	2	US-08-969-106-9	Sequence 9, Appl
32	40.2	2.6	2399	2	US-09-338-125-9	Sequence 9, Appl
33	40	2.6	293	4	US-09-313-294A-604	Sequence 604, App
34	40	2.6	762	4	US-09-252-991A-11986	Sequence 11986, A
35	40	2.6	1125	4	US-09-252-991A-11986	Sequence 11986, A
36	39.6	2.6	954	3	US-08-680-506-5	Sequence 5, Appl
37	39.6	2.6	1447	3	US-08-680-506-8	Sequence 8, Appl
38	39.6	2.6	1587	3	US-08-680-506-6	Sequence 6, Appl
39	39.6	2.6	2233	3	US-08-680-506-4	Sequence 4, Appl
40	39.6	2.6	4184	3	US-08-785-310A-4	Sequence 4, Appl
41	39.6	2.6	5027	3	US-08-680-506-2	Sequence 2, Appl
42	39.6	2.6	47981	4	US-09-679-279-1	Sequence 1, Appl
43	39.4	2.6	675	3	US-08-998-416-723	Sequence 723, App
44	39.2	2.5	1596	3	US-09-655-270A-2	Sequence 2, Appl
45	39.2	2.5	1596	3	US-09-651-941-2	Sequence 2, Appl
46	39.2	2.5	1596	4	US-09-955-597-2	Sequence 2, Appl
47	39.2	2.5	12508	3	US-09-655-270A-1	Sequence 1, Appl
48	39.2	2.5	12523	3	US-09-651-941-1	Sequence 1, Appl
49	39.2	2.5	12523	4	US-09-955-597-1	Sequence 1, Appl
50	39	2.5	531	4	US-09-252-991A-12051	Sequence 12051, A
51	39	2.5	564	4	US-09-252-991A-11993	Sequence 11993, A
52	39	2.5	603	4	US-09-252-991A-11962	Sequence 11962, A
53	39	2.5	645	4	US-09-252-991A-14041	Sequence 14041, A
54	38.8	2.5	708	4	US-09-252-991A-13744	Sequence 13744, A
55	38.8	2.5	804	4	US-09-252-991A-13605	Sequence 13605, A
56	38.8	2.5	984	4	US-09-252-991A-13494	Sequence 13494, A
57	38.8	2.5	1104	4	US-09-252-991A-1437	Sequence 1437, Ap
58	38.8	2.5	1275	4	US-09-252-991A-4071	Sequence 4071, Ap
59	38.8	2.5	2481	4	US-09-894-998A-35	Sequence 35, Appl
60	38.6	2.5	11958	3	US-09-134-246-8	Sequence 8, Appl
61	38.6	2.5	229354	4	US-09-765-400-64	Sequence 64, Appl
62	38.6	2.5	229354	4	US-09-705-400-64	Sequence 64, Appl
63	38.4	2.5	1683	4	US-09-724-797-21	Sequence 21, Appl
64	38.4	2.5	3524	4	US-09-077-940A-3	Sequence 3, Appl
65	38.4	2.5	3721	4	US-10-140-002-543	Sequence 543, App
66	38.4	2.5	4480	3	US-09-191-171-7	Sequence 7, Appl
67	38.4	2.5	4480	3	US-09-385-707-7	Sequence 7, Appl
68	38.2	2.5	6238	4	US-09-639-996C-6	Sequence 6, Appl
69	38.2	2.5	1260	4	US-09-252-991A-6674	Sequence 6674, Ap
70	38.2	2.5	2163	4	US-09-252-991A-6641	Sequence 6641, Ap
71	38.2	2.5	2889	1	US-08-289-112-1	Sequence 1, Appl
72	38.2	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl
73	38	2.5	702	1	US-08-458-568A-3	Sequence 3, Appl
74	38	2.5	1335	5	PCT-US91-06532-1	Sequence 1, Appl
75	38	2.5	2634	1	US-08-196-218-31	Sequence 31, Appl
76	38	2.5	2634	1	US-08-681-953-31	Sequence 31, Appl
77	38	2.5	2846	3	US-09-613-182-5	Sequence 5, Appl
78	38	2.5	12001	1	US-08-458-568A-11	Sequence 11, Appl
79	37.8	2.4	13442	3	US-09-105-537-30	Sequence 30, Appl
80	37.8	2.4	36778	3	US-09-105-537-5	Sequence 5, Appl
81	37.8	2.4	38506	3	US-09-320-876-19	Sequence 19, Appl
82	37.8	2.4	38506	4	US-09-141-908-1	Sequence 1, Appl
83	37.8	2.4	38506	4	US-09-657-440-19	Sequence 19, Appl
84	37.6	2.4	2560	3	US-09-305-640-1	Sequence 1, Appl
85	37.6	2.4	2868	4	US-10-140-002-119	Sequence 119, App
86	37.6	2.4	1706	4	US-09-174-528-383	Sequence 383, App
87	37.2	2.4	648	4	US-09-252-991A-9715	Sequence 9715, Ap
88	37.2	2.4	2150	2	US-08-318-837-1	Sequence 1, Appl
89	37.2	2.4	3395	3	US-09-428-711A-13	Sequence 13, Appl
90	37.2	2.4	5561	4	US-09-418-710-28	Sequence 28, Appl
91	37.2	2.4	5561	4	US-09-839-479-28	Sequence 28, Appl
92	37.2	2.4	5573	4	US-09-418-710-30	Sequence 30, Appl
93	37.2	2.4	5573	4	US-09-839-479-30	Sequence 30, Appl
94	37.2	2.4	6530	2	US-08-146-930-1	Sequence 1, Appl
95	37.2	2.4	6530	3	US-08-458-240-1	Sequence 1, Appl
96	37.2	2.4	6530	5	PCT-US93-03993-1	Sequence 1, Appl
97	37	2.4	516	3	US-09-183-861-34	Sequence 34, Appl
98	37	2.4	516	3	US-09-022-765-34	Sequence 34, Appl
99	37	2.4	516	4	US-09-551-974A-44	Sequence 44, Appl
100	37	2.4	516	4	US-09-555-501A-34	Sequence 34, Appl

C 101	37	2.4	516	4	US-09-639-206A-34	Sequence 34, Appl
C 102	37	2.4	516	4	US-09-874-923-34	Sequence 34, Appl
C 103	37	2.4	582	4	US-09-874-923-114	Sequence 114, Appl
C 104	37	2.4	591	4	US-09-252-991A-7544	Sequence 7544, Ap
C 105	37	2.4	648	4	US-09-252-991A-7795	Sequence 7795, Ap
C 106	37	2.4	786	4	US-09-252-991A-7871	Sequence 7871, Ap
C 107	37	2.4	1008	4	US-09-252-991A-7103	Sequence 7103, Ap
C 108	37	2.4	1347	4	US-09-252-991A-7693	Sequence 7693, Ap
C 109	37	2.4	7065	4	US-09-874-923-115	Sequence 115, App
C 110	37	2.4	10023	4	US-09-252-991A-6997	Sequence 6997, Ap
C 111	37	2.4	35100	2	US-08-770-379-17	Sequence 17, Appl
C 112	37	2.4	35100	3	US-08-757-669A-17	Sequence 17, Appl
C 113	37	2.4	35100	3	US-09-230-371A-17	Sequence 17, Appl
C 114	36.8	2.4	355	3	US-09-043-303-1	Sequence 1, Appl1
C 115	36.8	2.4	420	4	US-09-513-999C-31936	Sequence 31936, A
C 116	36.8	2.4	597	3	US-08-122-458D-17	Sequence 17, Appl
C 117	36.8	2.4	623	3	US-09-043-303-5	Sequence 5, Appl1
C 118	36.8	2.4	768	4	US-09-252-991A-6162	Sequence 6162, Ap
C 119	36.8	2.4	1034	4	US-09-023-655-1500	Sequence 1500, Ap
C 120	36.8	2.4	1057	1	US-08-147-784-1	Sequence 1, Appl1
C 121	36.8	2.4	1057	3	US-08-195-967-1	Sequence 1, Appl1
C 122	36.8	2.4	1057	3	US-08-472-940-1	Sequence 1, Appl1
C 123	36.8	2.4	1058	4	US-09-880-939-1	Sequence 1, Appl1
C 124	36.8	2.4	1058	4	US-09-804-200-1	Sequence 1, Appl1
C 125	36.8	2.4	1146	4	US-09-252-991A-6005	Sequence 6005, Ap
C 126	36.8	2.4	1827	4	US-09-252-991A-6085	Sequence 6085, Ap
C 127	36.8	2.4	1839	3	US-08-122-458D-10	Sequence 10, Appl
C 128	36.8	2.4	2258	4	US-09-016-434-1415	Sequence 1415, Ap
C 129	36.8	2.4	2561	4	US-09-616-289-48	Sequence 48, Appl
C 130	36.8	2.4	5117	4	US-09-774-528-438	Sequence 438, Ap
C 131	36.6	2.4	1334	4	US-09-149-476-217	Sequence 217, App
C 132	36.6	2.4	1407	4	US-09-252-991A-11113	Sequence 11113, A
C 133	36.6	2.4	1416	4	US-09-252-991A-11168	Sequence 11168, A
C 134	36.6	2.4	1590	4	US-09-252-991A-11088	Sequence 11088, A
C 135	36.6	2.4	1771	2	US-08-533-669A-7	Sequence 7, Appl1
C 136	36.6	2.4	1771	2	US-08-511-872-1	Sequence 1, Appl1
C 137	36.6	2.4	1771	3	US-09-183-861-7	Sequence 7, Appl1
C 138	36.6	2.4	1771	3	US-09-022-765-7	Sequence 7, Appl1
C 139	36.6	2.4	1771	4	US-09-551-974A-7	Sequence 7, Appl1
C 140	36.6	2.4	1771	4	US-09-565-501A-7	Sequence 7, Appl1
C 141	36.6	2.4	1771	4	US-09-639-206A-7	Sequence 7, Appl1
C 142	36.6	2.4	1771	4	US-09-874-923-7	Sequence 7, Appl1
C 143	36.6	2.4	1771	4	US-08-798-841-7	Sequence 7, Appl1
C 144	36.6	2.4	2828	4	US-09-016-434-1458	Sequence 1458, Ap
C 145	36.6	2.4	2980	4	US-09-266-225D-11	Sequence 11, Appl
C 146	36.6	2.4	7560	4	US-09-754-112A-2	Sequence 2, Appl1
C 147	36.6	2.4	12738	4	US-09-754-112A-1	Sequence 1, Appl1
C 148	36.4	2.4	540	4	US-09-252-991A-151	Sequence 151, App
C 149	36.4	2.4	666	4	US-09-252-991A-2156	Sequence 2156, Ap
C 150	36.4	2.4	897	4	US-09-252-991A-2083	Sequence 2083, Ap
C 151	36.4	2.4	1107	4	US-09-252-991A-1880	Sequence 1880, Ap
C 152	36.4	2.4	1248	4	US-09-252-991A-4845	Sequence 4845, Ap
C 153	36.4	2.4	1350	4	US-09-891-053-11	Sequence 11, Appl
C 154	36.4	2.4	1620	4	US-09-252-991A-4892	Sequence 4892, Ap
C 155	36.4	2.4	1632	4	US-09-252-991A-157	Sequence 157, App
C 156	36.4	2.4	2335	4	US-09-492-361-1	Sequence 1, Appl1
C 157	36.4	2.4	2689	2	US-08-985-090-1	Sequence 1, Appl1
C 158	36.4	2.4	2689	3	US-09-165-543-1	Sequence 3, Appl1
C 159	36.4	2.4	3889	4	US-09-484-970B-39	Sequence 39, Appl
C 160	36.2	2.3	761	4	US-09-799-451-769	Sequence 769, App
C 161	36.2	2.3	1482	4	US-09-252-991A-8172	Sequence 8172, Ap
C 162	36.2	2.3	1645	4	US-09-919-172-1	Sequence 1, Appl1
C 163	36.2	2.3	3096	4	US-09-544-398B-5	Sequence 5, Appl1
C 164	36.2	2.3	3096	4	US-09-543-771-5	Sequence 5, Appl1
C 165	36.2	2.3	4481	3	US-09-041-886-18	Sequence 18, Appl
C 166	36.2	2.3	4481	4	US-09-648-281-1	Sequence 1, Appl1
C 167	36.2	2.3	4481	4	US-09-707-919A-20	Sequence 20, Appl1
C 168	36.2	2.3	4481	4	US-09-083-268-2	Sequence 2, Appl1
C 169	36	2.3	369	4	US-09-313-294A-2951	Sequence 2951, Ap
C 170	36	2.3	364	4	US-09-621-976-17202	Sequence 17202, A
C 171	36	2.3	505	4	US-09-621-976-15639	Sequence 15639, A
C 172	36	2.3	1098	4	US-09-252-991A-14868	Sequence 14868, A
C 173	36	2.3	1194	4	US-09-252-991A-13276	Sequence 13276, A

174	36	2.3	1305	4	US-09-252-991A-12685	Sequence 12685, A
C 175	36	2.3	1473	4	US-09-252-991A-14632	Sequence 14632, A
176	36	2.3	1770	4	US-08-241-99A-25	Sequence 25, Appl
177	36	2.3	1770	4	US-09-635-133-2	Sequence 2, Appl1
178	36	2.3	1884	4	US-09-252-991A-15125	Sequence 15125, A
C 179	36	2.3	2019	1	US-08-254-357-3	Sequence 3, Appl1
C 180	36	2.3	2064	1	US-08-343-428-1	Sequence 1, Appl1
C 181	36	2.3	2768	4	US-09-380-777-1	Sequence 1, Appl1
C 182	36	2.3	4983	1	US-08-472-358-1	Sequence 1, Appl1
C 183	36	2.3	4983	5	PCT-US92-05766A-1	Sequence 1, Appl1
C 184	36	2.3	4984	1	US-08-687-806-1	Sequence 1, Appl1
C 185	36	2.3	4984	4	US-09-635-132-1	Sequence 1, Appl1
C 186	36	2.3	20235	1	US-07-642-734C-3	Sequence 3, Appl1
C 187	36	2.3	20235	3	US-08-439-090A-3	Sequence 3, Appl1
C 188	35.8	2.3	337	4	US-09-513-999C-1811	Sequence 1811, Ap
C 189	35.8	2.3	666	1	US-08-018-977C-2	Sequence 2, Appl1
C 190	35.8	2.3	1026	4	US-09-252-991A-9554	Sequence 9554, Ap
C 191	35.8	2.3	1399	3	US-09-188-930-198	Sequence 198, App
C 192	35.8	2.3	1399	4	US-09-312-283C-198	Sequence 198, App
C 193	35.8	2.3	1611	4	US-09-252-991A-9509	Sequence 9509, Ap
C 194	35.8	2.3	1741	1	US-08-565-655-5	Sequence 5, Appl1
C 195	35.8	2.3	2174	4	US-09-613-444-1	Sequence 1, Appl1
C 196	35.8	2.3	2511	4	US-09-919-497-36	Sequence 36, Appl
C 197	35.6	2.3	459	4	US-09-707-919A-13	Sequence 13, Appl1
C 198	35.6	2.3	516	4	US-09-083-268-1	Sequence 1, Appl1
C 199	35.6	2.3	663	4	US-09-252-991A-5078	Sequence 5078, Ap
C 200	35.6	2.3	720	4	US-09-252-991A-7723	Sequence 7723, Ap

## ALIGNMENTS

```
RESULT 1
US-08-913-832A-1
; Sequence 1, Application US/08913832A
; Patent No. 6329517
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0030-999
; CURRENT APPLICATION NUMBER: US/08/913,832A
; PRIOR APPLICATION NUMBER: 1998-01-12
; PRIOR FILING DATE: PCT/DE96/00444
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(5736)
US-08-913-832A-1
```

Query Match 3.9%; Score 60; DB 3; Length 6328;

Best Local Similarity 61.5%; Pred No. 3.2e-05;

Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY	532	AGGACGAGTGTCCGCTGTGTGCGGACGGCGGGAGCTCATCTGTGTGACGCGTCCCTC	591
DB	1109	AGGACGATTTGGCAGGAGTGTGCGGACGAGGGGCGTGAGTATCTGTGTATCTGTACCC	1168
QY	592	GGGCTTCACCTGGCTGCTGTCTCCCTCCGCTCCGGAGATCCCCCACTGGAGACTTGA	651
DB	1169	GTGCTTACCACTGATGTCTGCTCGGATCCCGACATGAGAAAGCTCCGAGGCGAAGTGA	1228
QY	652	GGTGTCCAGCTGCGGACGAGCAAGTCCAGGAGG	687
DB	1229	GCTGCCACACTGCGAGAAAGGAGCATTCAGTGGG	1264

RESULT 2  
US-09-249-181A-1  
; Sequence 1, Application US/09249181A  
; Patent No. 6440679  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred  
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
; FILE REFERENCE: 8484-0059-999  
; CURRENT APPLICATION NUMBER: US/09/249,181A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 08/913,832  
; PRIOR FILING DATE: 1998-01-12  
; PRIOR APPLICATION NUMBER: PCT/DE96/00444  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 6328  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(5736)  
US-09-249-181A-1

Query Match 3.9%; Score 60; DB 4; Length 6328;  
Best Local Similarity 61.5%; Pred. No. 3.2e-05;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTCGGGAGCGGGAGAGCTCATCTCTGTGACGGCTGCCCTC 591  
DB 1109 AGGACTATTGCGAGGTGTGCGACCAAGCGGTGAGTCACTGTGTGATACCTGTCCCC 1168  
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTGGA 651  
DB 1169 GTGCTTACCACTGCTGTGCTGCTGATCCCACTGAGAGAGCTCCCGAGGGCAAGTGA 1228  
QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGG 687  
DB 1229 GCTGCCACACTGCGAAGAGGAGCATCCAGTGGG 1264

RESULT 3  
US-09-158-707-1  
; Sequence 1, Application US/09158707  
; Patent No. 6500923  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred  
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
; FILE REFERENCE: 8484-0043-999  
; CURRENT APPLICATION NUMBER: US/09/158,707  
; CURRENT FILING DATE: 1998-09-22  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 6328  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(5736)  
US-09-158-707-1

Query Match 3.9%; Score 60; DB 4; Length 6328;  
Best Local Similarity 61.5%; Pred. No. 3.2e-05;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTCGGGAGCGGGAGAGCTCATCTCTGTGACGGCTGCCCTC 591  
DB 1109 AGGACTATTGCGAGGTGTGCGACCAAGCGGTGAGTCACTGTGTGATACCTGTCCCC 1168

QY 592 GGGCTTCCACCTGGCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTGGA 651  
DB 1169 GTGCTTACCACTGCTGTGCTGCTGATCCGACATGAGAGGCTCCGAGGCAAGTGA 1228  
QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGG 687  
DB 1229 GCTGCCACACTGCGAAGAGGAGCATCCAGTGGG 1264

RESULT 4  
US-09-620-312D-325  
; Sequence 325, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonhong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhilwei  
; APPLICANT: John Thilloghaet  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pc\_FL\_genes Version 1.0  
; SEQ ID NO 325  
; LENGTH: 6475  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (219)...(5957)  
US-09-620-312D-325

Query Match 3.9%; Score 60; DB 4; Length 6475;  
Best Local Similarity 61.5%; Pred. No. 3.3e-05;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTCGGGAGCGGGAGAGCTCATCTGTGACGGCTGCCCTC 591  
DB 1327 AGGACTATTGCGAGGTGTGCGACCAAGCGGTGAGTCACTGTGTGATACCTGTCCCC 1386  
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTGGA 651  
DB 1387 GTGCTTACCACTGCTGTGCTGCTGATCCGACATGAGAGGCTCCCGAGGGCAAGTGA 1446  
QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGG 687  
DB 1447 GCTGCCACACTGCGAAGAGGAGCATCCAGTGGG 1482

RESULT 5  
US-07-945-295-1  
; Sequence 1, Application US/07945295  
; Patent No. 6610823  
; GENERAL INFORMATION:

```

; APPLICANT: Targoff, Ira N.
; APPLICANT: Ge Qun
; TITLE OF INVENTION: Antigens Associated with Polymyositis
; TITLE OF INVENTION: and with Dermatomyositis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Ste. 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,295
; FILING DATE: 19920909
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR# 120C1P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Blood
; IMMEDIATE SOURCE:
; LIBRARY: Human thymocyte lambda gt11
; CLONE: L1
; US-07-945-295-1

Query Match          3.9%; Score 59.8; DB 4; Length 1053;
Best Local Similarity 58.9%; Pred. No. 2e-05;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTCGCGGAGCGGGGAGCTCATCTGCTGAGACGGCTGCCCTC 591
    |||||
DB 248 AGGATTACTGTGAGGTGTGCGACGAGGTGGGAAATTATCTGTGTGACACTTGCCCTC 307
    |||||

QY 592 GGGCCTTCACCTGAGCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    |||||
DB 308 GTGCTTACCACTCGTCTGCTGTGATCTGAGCTTACCGGGGTCCAGAGGGGCAATGGA 367
    |||||

QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGGTGACAGCCCCGGGAGAGGA 706
    |||||
DB 368 GCTGCCCTCACTGTGAGAGGAGGGGCTCAGTGGAGGCCAAGAGGAAGA 422
    |||||

RESULT 6
PCT-US91-06418-2
; Sequence 2, Application PC/TUS9106418
; GENERAL INFORMATION:
; APPLICANT: Oklahoma Medical Research, Foundation, et al
; TITLE OF INVENTION: Antigens Associated with Polymyositis
; TITLE OF INVENTION: and with Dermatomyositis
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
```

```

; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06418
; FILING DATE: 19910905
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/579023
; FILING DATE: 09-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR#120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Sera
; IMMEDIATE SOURCE:
; CLONE: L1
; PCT-US91-06418-2

Query Match          3.9%; Score 59.8; DB 5; Length 1053;
Best Local Similarity 58.9%; Pred. No. 2e-05;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTCGCGGAGCGGGGAGCTCATCTGCTGAGACGGCTGCCCTC 591
    |||||
DB 248 AGGATTACTGTGAGGTGTGCGACGAGGTGGGAAATTATCTGTGTGACACTTGCCCTC 307
    |||||

QY 592 GGGCCTTCACCTGAGCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    |||||
DB 308 GTGCTTACCACTCGTCTGCTGTGATCTGAGCTTGAACCGGGCTCCAGAGGGGCAATGGA 367
    |||||

QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGGTGACAGCCCCGGGAGAGGA 706
    |||||
DB 368 GCTGCCCTCACTGTGAGAGGAGGGGCTCAGTGGAGGCCAAGAGGAAGA 422
    |||||

RESULT 7
US-08-942-686-1
; Sequence 1, Application US/08942686
; Patent No. 6183988
; GENERAL INFORMATION:
; APPLICANT: Bloch, Donald B.
; TITLE OF INVENTION: LEUCOCYTE-SPECIFIC PROTEIN AND GENE, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
```



ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,686  
FILING DATE: 02-OCT-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/027,347  
FILING DATE: 02-OCT-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: GOLDSTEIN, JORGE A.  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 0609.4330001/JAG/BJD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2905 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 107..2365  
US-08-942-686-1

Query Match 3.1%; Score 47.2; DB 3; Length 2905;  
Best Local Similarity 64.8%; Pred. No. 0.036;  
Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 515 CCTGTGCCAGAAATGAGAGACAGTGTGCTGTGTGCGAGCGGGGAGCTCATCTG 574  
DB 1816 CCTGTGTATGAGAACTGATGAGTGTGAGGTGCTCCGGACGGAGGAGCTGTCTG 1875  
QY 575 CTGTGACGGCTGCTCGGAGCTTCCACCTGCGCTGCTGCTCCCTCC 622  
DB 1876 TTGCGACACTTGTTCAGAGTCTTCCATGAGGACTGTCAATCCGCC 1923

RESULT 8  
US-09-976-594-1064/c  
Sequence 1064, Application US/09976594  
Patent No. 6673549  
GENERAL INFORMATION:  
APPLICANT: Furness, Michael  
APPLICANT: Buchbinder, Jenny  
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
FILE REFERENCE: PA-0041 US  
CURRENT APPLICATION NUMBER: US/09/976,594  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 60/240,409  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 1143  
SOFTWARE: PERL Program  
SEQ ID NO 1064  
LENGTH: 2327  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Inocyte ID No. 6673549 210133.1  
US-09-976-594-1064

Query Match 2.9%; Score 44.6; DB 4; Length 2327;  
Best Local Similarity 50.7%; Pred. No. 0.15;  
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 568 TCATCTGCTGTGACGAGCTGCTCGGAGCTTCCACCTGCGCTGTGCTCCCTCCGCTCC 627  
DB 417 TCGTGGCTGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 358  
QY 628 GGGAGATCCCAAGTGGAGCTGAGAGTGTCTGAGCTGCTGAGGCAACAGTCCAGAGAG 687  
DB 357 GCGAGGCGAGCGGGGCCCCCGGAGAGCGGCGGTCCCAACAGCTCTTCGAG 298  
QY 688 TCGAGCCCCGGGCAAGAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747  
DB 297 CCTGCTGCTGGGAGAGCGGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 238  
QY 748 CGAGGCTTAAAGTGGCGGAGAGAGAGTAAAG 778  
DB 237 TCGTCTTCATCCAGATAGAGAGAGACTAG 207

RESULT 9  
US-08-340-203A-1/c  
Sequence 1, Application US/08340203A  
Patent No. 5756668  
GENERAL INFORMATION:  
APPLICANT: Baylin, Stephen B.  
TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,203A  
FILING DATE: 15-NOV-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: P-38,347  
REFERENCE/DOCKET NUMBER: 07265/039001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4616 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: HIC-1 polynucleotide  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..4616  
US-08-340-203A-1

Query Match 2.9%; Score 44.6; DB 1; Length 4616;  
Best Local Similarity 50.7%; Pred. No. 0.19;  
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 568 TCATCTGCTGTGACGAGCTGCTCGGAGCTTCCACCTGCGCTGTGCTCCCTCCGCTCC 627  
DB 2211 TCGTGGCTGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 2152  
QY 628 GGGAGATCCCAAGTGGAGAGCTGAGAGTGTCTGAGCTGCTGAGGCAACAGTCCAGAGAG 687







```

Query Match 2.7%; Score 42.2; DB 4; Length 3096;
Best Local Similarity 61.3%; Pred. No. 0.64;
Matches 68; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 557 CGCGGGAGCTCATCTGCTGTGACGGCTGCGCTCGGGCTTGCACCTGAGCTGCTGTC 616
Db 2062 CGTGGGGAGATGATCTACTGTGACTTGTGCTGTGCTGTCCTTACACACGGCGAATTGCTATAT 2003

Qy 617 CCCTTCGCTCCGGAGATCCCACTGAGGAGCTGAGAGTGTCCAGCTGCT 667
Db 2002 ACCGCGCTGCTAAAGGATACCGCGGAGAGGAGTACCGACAGGTTGCAT 1952

RESULT 18
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007_00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.7%; Score 42.2; DB 3; Length 4403765;
Best Local Similarity 45.0%; Pred. No. 7.2;
Matches 203; Conservative 0; Mismatches 243; Indels 5; Gaps 1;

Qy 466 CTCTCAGAGAGGTGTGTCACCCCGACCCAGCTGATGAGGAGTCTTGGCTTGCCAGA 525
Db 2956329 CTGACCGCGGTGTGCGCGCTGTCGCGCTGTCGCGCTGTCGCGCTGTCGCGCGGAGCC 2956388

Qy 526 AGAATGAGAGCAGATGTGTCGCTGTGTGTCGAGACGGCGGAGACTCATCTGTGTGACGGCT 585
Db 2956389 ACCTTGCCCAACCGTGCGCGCTGTGCGCGGTGTCGCGCACTCTGTCGCCACCGGATCAG 2956448

Qy 586 GCCCTCGGAGCTTTCACCTGTGACCTGTGCTGTGCTTCCCTTCGCTTCGGAGATCCCACTGAGGA 645
Db 2956449 CCCCGCGGCTTCGCGCGCTGTCGCGCGGTGTGTCGCGCGCGCGCGCGCGGATCCAGTGC 2956508

Qy 646 CCTGAGAGTGTCTCAGCTGCTGCT-----GCAGGCAACATGCCAGAGGTGACGCCCGGCGC 700
Db 2956509 ACCAGCGCCCCCGACGCGCTGTGACGCGCGCGCACCGCAATGTGCTCTCCCGCGCGGT 2956568

Qy 701 AGAGAGCGCCCGCGCCCGACAGACCCCGTGAAGACCCCGCTCCCGCGGAGCTTAAGTGC 760
Db 2956569 GGTGGTACCGCGCGCGCGCGCGCGACGTTGCGCGCGGACACACGATGCGCGCGGATGCC 2956628

Qy 761 GCGCGGAGAGAGAGTGAAGAGTTCACCTCTGGGAAACCTTACCGGATGAGACAGACTCT 820
Db 2956629 ACCGGTGCAGACCGCGCGGACCCCGCGCACACACAGAGCCGATGAGACCGAGC 2956688

Qy 821 TGTCTACAAAGACACTGCGCGCTTCGCTTGTGACAGCCCGCTGCACGATGTGACTCTC 880
Db 2956689 CGCCCGCGCTTGGCGCGCGGACACCGCGCGCTGCACCATTCGCGCGTACCGCGGACAC 2956748

Qy 881 GGCGCTGACCCCTACTGTGTGTGGTCT 911

```

;/ TITLE OF INVENTION: A new method of identifying non-host plant disease  
;/ TITLE OF INVENTION: resistance genes  
;/ FILE REFERENCE: r gene patent  
;/ CURRENT APPLICATION NUMBER: US/09/387,286  
;/ EARLIER FILING DATE: 1999-08-31  
;/ EARLIER APPLICATION NUMBER: 60/098,402  
;/ EARLIER FILING DATE: 1998-08-31  
;/ NUMBER OF SEQ ID NOS: 66  
;/ SOFTWARE: Patent In Ver. 2.0  
;/ SEQ ID NO: 60  
;/ LENGTH: 3612  
;/ TYPE: DNA  
;/ ORGANISM: Nicotiana tabacum  
US-09-387-286-60

Query Match  
Best Local Similarity 66.7%; Score 42; DB 4; Length 3612;  
Pred. No. 0.76; Mismatches 30; Indels 0; Gaps 0;  
Matches 60; Conservative 0;

QY 521 CCAGAAAGATGAGAGAGAGTGTGCGGTGTCGAGAGCGGGAGCTCATCTGCTGTGA 580  
DB 1961 CATAACAGTATGATGATGTGTCATATGCGCGATGGGGAGAACTGATTGCTGTGA 2020  
QY 581 CGGCTGCGCTCGGCGCTTCACCTGCGCTG 610  
DB 2021 AGGTTGTCTCGGGCTTTCATGACAGCTTG 2050

RESULT 21  
PCT-US91-07635-5  
; Sequence 5, Application PC/TUS9107635  
; GENERAL INFORMATION:  
; APPLICANT: OPPERMANN, HERMANN  
; APPLICANT: OZKAYNAK, ENGIN  
; APPLICANT: RUBGER, DAVID C  
; APPLICANT: KUBERASAMPATH, THANGAVEL  
; TITLE OF INVENTION: OSTEOGENIC DEVICES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS: '  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07635  
; FILING DATE: 19911018  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDWARD R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: CRP-056PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1941 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HOMO SAPIENS  
; TISSUE TYPE: HIPPOCAMPUS

;/ FEATURE:  
;/ NAME/KEY: CDS  
;/ LOCATION: 507..1703  
;/ OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"  
;/ OTHER INFORMATION: /product= "HOP2"  
;/ OTHER INFORMATION:  
PCT-US91-07635-5

Query Match  
Best Local Similarity 51.6%; Score 41.6; DB 5; Length 1941;  
Pred. No. 0.77; Mismatches 89; Indels 0; Gaps 0;  
Matches 95; Conservative 0;

QY 566 GCTCATCTGCTGTGACGAGTGTGCTTCCGCTTCACCTGAGCTGCTGCTCCCTCCGCT 625  
DB 533 GCTCTCGGCTCTGGGGCTATGCGCGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 592  
QY 626 CCGGAGATCCCAAGTGGAGCTGGAGGTGCTTCAGCTGCTGACGAGCAAGTCCAGGA 685  
DB 593 CGGCTGTCCCGACGACGCTGTGGGGCGCGCGACCGGAGACGTGACGCGGAGATCTGGGC 652  
QY 686 GGTGACAGCCCGGGGAGAGAGCCCGCGCCGAGAGACCGGTGGAGACCCCGCTGCC 745  
DB 653 GGTGCTCGGGCTGCTGTGG 712  
QY 746 CCCG 749  
DB 713 CGCG 716

RESULT 22  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION NUMBER: 435  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMWU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

TOPOLOGY: 1linear  
IMMEDIATE SOURCE:  
CLONE: PTZpt-F18  
US-08-232-463-14

Query Match  
Best Local Similarity 3.1%; Pred. No. 1.3; Length 7218;  
Matches 12; Conservative 209; Mismatches 160; Indels 0; Gaps 0;

QY 225 GAGGAGGGTTCATGTTGGTGTACAGTTCGGGGCCCTGGAAAGCAGCAGCCTGCA 284  
DB 1426 RRR 1367  
QY 285 AAAAAAGGGTTTCTTCCCATAGGATGCCCCGGGGGGTCTCTTTCAGACCAAT 344  
DB 1366 RRR 1307  
QY 345 GGATGGGGAACAGTGTCAAGGAGCAATTTGAGCCCTGGCAGCATGGAGAGGAGCAG 404  
DB 1306 RRR 1247  
QY 405 AGACTGGGAGTTCAGTACCAAGATGCTGTGGGAGAGCTTTTGGAGAGAGTG 464  
DB 1246 RRR 1187  
QY 465 GCTCTCAGAGAGGTGCTGACCCAGCCAGCTGATGAGGGCTCTTCTGCTGCCAG 524  
DB 1186 RRR 1127  
QY 525 AAGATGAGAGCAAGTGTCCGTGTGTCGAGAGCGGGAGAGCTATGCTGTGACGGC 584  
DB 1126 RRR 1067  
QY 585 TGCCCTCGGGCTTTCACCTG 605  
DB 1066 ATCGCAAGCTCCCTGCACCTG 1046

## RESULT 23

US-09-312-762A-3  
Sequence 3, Application US/09312762A  
Patent No. 6552177  
GENERAL INFORMATION:  
APPLICANT: MIA HOROWITZ ET AL.  
TITLE OF INVENTION: EH DOMAIN CONTAINING GENES AND PROTEINS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
STREET: 2001 Jefferson Davis Highway, Suite 207  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Twinhead\* Slimnote-890TX  
OPERATING SYSTEM: MS DOS version 6.2,  
SOFTWARE: Word for Windows version 2.0 converted to  
SOFTWARE: an ASCII file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,762A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/026,898  
FILING DATE: 20 FEB 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33, 883  
REFERENCE/DOCKET NUMBER: 916/10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 972-3-562553

TELEFAX: 972-3-5625554  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14707  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: 1linear  
US-09-312-762A-3

Query Match  
Best Local Similarity 53.0%; Pred. No. 1.9; Length 14707;  
Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 980 GTGTACTACAGCGCGCTGCTTCCACTGAGCGCTGCACTTCCAGCCGAGCCTCCG 1039  
DB 10156 GTCTCTTCTTCTTACTGACAGCTTGAGGATTTCCACCCAGAGCAGATAGAGCT 10215  
QY 1040 GCCCGGAGCGGCTCCGCTGCAGATCTGCTCAGAGACGTGACCCCGCCCTGTGA 1099  
DB 10216 GACATCTTAGGGCTTGAGTTGCTTCTGCTGAGGAGGTATCTCCAGCTTGTCTGA 10275  
QY 1100 GGGGCTGAGCGCCCGCCGCGCCGCTGCGCCCTGCGCTGCGC 1145  
DB 10276 GAGGGGTCTGAATGTAAGCCCCCAGCCCGCCCGCCAGGTCAAC 10321

## RESULT 24

US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6284328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match  
Best Local Similarity 51.4%; Pred. No. 14;  
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 989 CTGGCCCGCTGCTTCCACTGCGCTGCACATTCCAGCGGCACTCCGCGCGGAGC 1048  
DB 1217931 CCGGCGCGCAAGCGCGCGGCGGCGGAGTCCGTGCGGCGCGCGCGCTCAC 1217872  
QY 1049 GGGGCTGCGCTGCAATCTGCTCAGAGACGTGACCCAGCCCTGTGAGAGGGGTGCT 1108  
DB 1217871 CGCCCTTACGCGCGCGCGCGCGCGCGCGCTACCGCGGTGCGCGGTGCCACCA 1217812  
QY 1109 GGGCCCGAGCCCGCGCGCTGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG 1168  
DB 1217811 GCAGCCCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGC 1217752  
QY 1169 GCGCG 1173  
DB 1217751 CGCGG 1217747

## RESULT 25

US-09-621-976-17202





```
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE K
APPLICANT: GROSS, MITCHELL S
APPLICANT: HURLE, MARK R
TITLE OF INVENTION: HUMAN REQUIEM
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAYNER & PRESTITA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/881,857
FILING DATE: 24-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/021,229
FILING DATE: 26-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESTITA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1556 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-881-857-1

Query Match
Best Local Similarity 59.0%; Score 40.2; DB 2; Length 1556;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 554 GGACGGCGGGAGCTCATCTGTGTGACGCGCTGCTCGGGCCTTCACCTGGCCTGCT 613
DB 1073 GAATGACGACCACTGCTCTTCTGTGATGATGATGATGATGATGATGATGATGAT 1132

QY 614 GTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGAGAGTGTCTCAGCTGCTGCA 670
DB 1133 CACCCGCTCATGTGTGACGCCCTGAGAGAGTTGAGAGCTGCACTGTGTCTGGA 1189

RESULT 29
US-09-233-342A-1
Sequence 1, Application US/09233342A
Patent No. 6207803
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE K.
APPLICANT: GROSS, MITCHELL S.
APPLICANT: HURLE, MARK ROBERT
TITLE OF INVENTION: HUMAN REQUIEM
FILE REFERENCE: ATG-50013-1
CURRENT APPLICATION NUMBER: US/09/233,342A
CURRENT FILING DATE: 1999-01-19
PRIOR APPLICATION NUMBER: 08/881,857
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: 60/021,299
PRIOR FILING DATE: 1996-06-26
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Fastseq for Windows Version 3.0
```

```
SEQ ID NO 1
LENGTH: 1556
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-233-342A-1

Query Match
Best Local Similarity 59.0%; Score 40.2; DB 3; Length 1556;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 554 GGACGGCGGGAGCTCATCTGTGTGACGCGCTGCTCGGGCCTTCACCTGGCCTGCT 613
DB 1073 GAATGACGACCACTGCTCTTCTGTGATGATGATGATGATGATGATGATGATGAT 1132

QY 614 GTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGAGAGTGTCTCAGCTGCTGCA 670
DB 1133 CACCCGCTCATGTGTGACGCCCTGAGAGAGTTGAGAGCTGCACTGTGTCTGGA 1189

RESULT 30
US-09-620-312D-389/c
Sequence 389, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CI2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_FL_genes Version 1.0
SEQ ID NO 389
LENGTH: 1682
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (25) .. (1197)
US-09-620-312D-389

Query Match
Best Local Similarity 49.3%; Score 40.2; DB 4; Length 1682;
Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 964 GTACGAGAGCTGTGGGAGTACTGACGCGCTGCTTCACCTGGCCTGCACTTCC 1023
DB 217 GCACGAGAGACTGGCCCGCATTCGCGCTGCTGCGCCGACGAGCGCTCTCTCCG 158

QY 1024 CAGCGGACCTTCCGCGCGCGGAGCGAGCGCTGCGCTGACAGATCTGCTCAGAGACGTGA 1083
DB 157 CCCGCTCATGTGTGACGACGCTCGCTACACCTTGGGCGCGCGCGAGGACCGGGA 98
```



Query Match 2.6%; Score 40; DB 4; Length 293;  
Best Local Similarity 60.6%; Pred. No. 1;  
Matches 86; Conservative 0; Mismatches 50; Indels 6; Gaps 1;

OY 526 AGAATGAGACGAGGTGCGGTGTCGGAGCGCGGAGGCTCATCTGTGAGCGCT 585  
DB 18 AAAAGATGATATACCTGTGATCATGTGGCATGTGTGTAACTGATATCTGTGACATTT 77  
OY 586 GCCCTGGGCGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGGA 645  
DB 78 GTCTGCTGATGATACCATCAAGCTGTGCTTGTG-----CCAGATATTTCCAGATGGCA 131  
OY 646 CCTGAGAGGTCTCCAGCTGCTT 667  
DB 132 ACTGTACTGCTCTGTAGCTGCTT 153

RESULT 34  
US-09-252-991A-11986  
Sequence 11986, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11986  
LENGTH: 762  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11986

Query Match 2.6%; Score 40; DB 4; Length 762;  
Best Local Similarity 50.5%; Pred. No. 1.4;  
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY 1096 TGGAGGGGCTGCTGCGCCCGCCGCGCTGCGCCCTGCGCCCTGCGCAAGATGACA 1155  
DB 386 TGGTGGAGGTGTGATGCGCGGAGAAAGACATTCGCGCGCGCTCCCTAGCAAGTGA 445  
OY 1156 CTGCAAGTACAGAGCCCGCTGTGCAAGGATGACCTGAGTCCCTTGTGACGAGCACA 1215  
DB 446 CACCGAAGGTGAAGCGGTGTGCGCGCGCGGAGGAGTCTATCCACGCGGCACA 505  
OY 1216 CTTGATGAGCATCTGCGAGTGGGCGCATCGAGCATGAGCCCGCGCGGCCCTTCC 1275  
DB 506 CCATGACAGCATCTTCAAGGCGCTGCTGTAGCATGTGTGCTCCGAGCATCAGATG 565  
OY 1276 CCTCTGACCCC 1287  
DB 566 CGTCTCGCCC 577

RESULT 35  
US-09-252-991A-11956  
Sequence 11956, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11956  
LENGTH: 1125  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11956

Query Match 2.6%; Score 40; DB 4; Length 1125;  
Best Local Similarity 50.5%; Pred. No. 1.6;  
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY 1096 TGGAGGGGCTGCTGCGCCCGCCGCGCTGCGCCCTGCGCCCTGCGCAAGATGACA 1155  
DB 576 TGGTGGAGGTGTGATGCGCGGAGAAAGACATTCGCGCGCGCTCCCTAGCAAGTGA 635  
OY 1156 CTGCAAGTACAGAGCCCGCTGTGCAAGGATGACCTGAGTCCCTTGTGACGAGCACA 1215  
DB 636 CACCGAAGGTGAAGCGGTGTGCGCGCGCGGAGGAGTCTATCCACGCGGCACA 695  
OY 1216 CTTGATGAGCATCTGCGAGTGGGCGCATCGAGCATGAGCCCGCGCGGCCCTTCC 1275  
DB 696 CCATGACAGCATCTTCAAGGCGCTGCTGTAGCATGTGTGCTCCGAGCATCAGATG 755  
OY 1276 CCTCTGACCCC 1287  
DB 756 CGTCTCGCCC 767

RESULT 36  
US-08-680-506-5  
Sequence 5, Application US/08680506C  
Patent No. 6008013  
GENERAL INFORMATION:  
APPLICANT: Reynolds, Paul R.  
TITLE OF INVENTION: CHONDROCYTE PROTEINS  
FILE REFERENCE: 176/60091  
CURRENT APPLICATION NUMBER: US/08/680,506C  
CURRENT FILING DATE: 1996-07-08  
EARLIER APPLICATION NUMBER: 60/021,672  
EARLIER FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 5  
LENGTH: 954  
TYPE: DNA  
ORGANISM: Gallus gallus  
US-08-680-506-5

Query Match 2.6%; Score 39.6; DB 3; Length 954;  
Best Local Similarity 44.9%; Pred. No. 1.9;  
Matches 150; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

OY 985 CTCAGTGGCGCGCTGCTTCACTGCGCTGCACTTCCAGCGACCTTCCGCGCCG 1044  
DB 5 CGCGGTGCGCGCTGCTGCTGCGCTTCTGTGCGGTGCGAGCGCGCGGCTGG 64  
OY 1045 GGAAGGGGCTGCGTGCAGATCTGCTGAGAGAGTACCCGACCCCTGAGGAGGGG 1104  
DB 65 GCGGCTACTTGGGACCAAGTCCGCTACAGAGAGTAAACCGACCTGCGGAGAGCC 124  
OY 1105 TGGTGGCCCGACGCGCGCGCTGCGCCCTGCGCTGCGCAAGATGACATGCGCATG 1164  
DB 125 CGCTGTCTCTGGGCGGACGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCG 184  
OY 1165 ACGAGCGGCTGTGACAGGAGTGACTGAGTCCCTTGTGAGCAGCAGACCTTGTGATG 1224  
DB 185 AGCTCGCGCGGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 244  
OY 1225 GCATCTGAGTGGGCGATTCAGAGCATGAGCGCGCGCGCGCGCGCGCGCGCGCG 1284



Db 325 TGCCGAGCTGCACGGCCCTCCGCGCGCCCGCCCGCTCTGCTCCCGCCCGCCG 384  
QY 1285 CCCAGATGGCCGGGACATGCAGCTCTGATGAGAG 1318  
Db 385 CGCTGGCCGCTGGCCGATGTGTACGAGAGAG 418

RESULT 40  
US-08-785-310A-4/c  
Sequence 4, Application US/08785310A  
Patent No. 5840532  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L.  
APPLICANT: Russell, David W.  
TITLE OF INVENTION: Neuronal PAS Domain Protein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,310A  
FILING DATE: 21-JAN-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UTSD:1226  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4184 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-785-310A-4

Query Match 2.6%; Score 39.6; DB 2; Length 4184;  
Best Local Similarity 50.4%; Pred. No. 3.1;  
Matches 122; Conservative 0; Mismatches 119; Indels 1; Gaps 1;  
QY 276 CAGCTGCAAGAAACCGGGTTTCTCCCAATAGGAGTGGCCCGGGGGGTGTCTGTGG 335  
Db 251 CACTCTCCAGGGCTCCCGGGCTCCCTACGCTTCGGGATGCTGCGGTGGGCTCTTGAAG 192  
QY 336 AGACCAAGATGATGGGAAACAGT-GGTACAGGGCAGAAATTCAGGCTTCGACAGATGG 394  
Db 191 GCGAGGGGCAATCTTGAAGCTTGTGTCGGGCAAGTACCCCTTACGACCTCCG 132  
QY 395 AGCAGGGCAGAGATCGGGAGTTCAGTACCCAGAAATGTGTGGGGAGAGCTTTTGG 454  
Db 131 AGGATCCCGGAATCCGATGTTGGGAGCTTACGTCTCTTGGCAGAGCTCCG 72  
QY 455 GAAGAGATGCTCTCAGAGAGGTGTGCACCCAGCCAGATGTGATGGGCGTCTCTTG 514  
Db 71 GAGCCCAACGCTCTGTCTGCTGAGCGGTTGGCAGAGACCCGAAGCCAGGGCGGCTGTG 12  
QY 515 CC 516  
Db 11 CC 10

RESULT 41  
US-08-680-506-2  
Sequence 2, Application US/08680506C  
Patent No. 6008013  
GENERAL INFORMATION:  
APPLICANT: Reynolds, Paul R.  
TITLE OF INVENTION: CHONDROCYTE PROTEINS  
FILE REFERENCE: 176/60091  
CURRENT APPLICATION NUMBER: US/08/680,506C  
CURRENT FILING DATE: 1996-07-08  
EARLIER APPLICATION NUMBER: 60/021,672  
EARLIER FILING DATE: 1996-07-05  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
TYPE: DNA  
LENGTH: 5027  
ORGANISM: Gallus gallus  
US-08-680-506-2

Query Match 2.6%; Score 39.6; DB 3; Length 5027;  
Best Local Similarity 44.9%; Pred. No. 3.3;  
Matches 150; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 985 CTCACCTGCGCGCTGCTTCCTCCACTGCGGCTGCCACTTCCAGCCGCGACTTCCCGCCG 1044  
Db 85 CCGCGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144  
QY 1045 GAGCGGCGCTGCGCTGCAAGATCTGCTCAGAGAGTGAAGCCAGCCCGCTGTGAGAGGG 1104  
Db 145 GCGGCTACTTGGGACACAGTCCGCTTACGAGAGAGTGAAGCCAGCTGCGGAGAGACC 204  
QY 1105 TCGTGGCCCCAGCCCGCCGCTGCGCCCTGCGGCTGCGCAAGATGCACTGCCAGTC 1164  
Db 205 CCGTGCCTCGGGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 264  
QY 1165 AGCAGCGCGCTTGCACAGGATGACCTGAGATCCCTTTGAGCGAGCACACTTGTGATG 1224  
Db 265 ACTCGCGCGCTGCTGCTGCGCACAGGACCCGCTTACCCAGCGCGCGCAAAATCCGCGCC 324  
QY 1225 GCATCCTGAGTGGGCAATCCAGAGATGAGCCGTCGGCGGCGCCCTTCCCTCTGTAC 1284  
Db 325 TGCCGAGCTGCACGGCCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384  
QY 1285 CCCAGATGGCCGGGACATGACAGCTCTGATGAGAG 1318  
Db 385 CGCTGGCCGCTGGCCGATGTGTACGAGAGAG 418

RESULT 42  
US-09-679-279-1/c  
Sequence 1, Application US/09679279  
Patent No. 6524841  
GENERAL INFORMATION:  
APPLICANT: McDaniel, Robert  
APPLICANT: Volchegursky, Yanina  
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic  
FILE REFERENCE: 300623004700  
CURRENT APPLICATION NUMBER: US/09/679,279  
CURRENT FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/158,305  
PRIOR FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 60/190,024  
PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 47981  
TYPE: DNA  
ORGANISM: Micromonospora megalomicea  
FEATURE:

```
NAME/KEY: CDS
LOCATION: (1)....(144)
OTHER INFORMATION: megBVI(megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (928)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megC(megY), mycarosyl acyltransferase, mycarose O-acyltransferase
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDII, deoxysugar transferase (eryCI, DnrJ homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (5822)...(6595)
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmnu hc
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmnu homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc_feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc_feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc_feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megAI, KS1
NAME/KEY: misc_feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megAI, ATI
NAME/KEY: misc_feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megAI, KRI
NAME/KEY: misc_feature
LOCATION: (17947)...(18207)
OTHER INFORMATION: megAI, ACP1
NAME/KEY: misc_feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megAI, KS2
NAME/KEY: misc_feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megAI, AT2
NAME/KEY: misc_feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megAI, KR2
NAME/KEY: misc_feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megAI, ACP2
NAME/KEY: CDS
LOCATION: (22867)...(33555)
OTHER INFORMATION: megAI, SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc_feature
LOCATION: (32957)...(42427)
OTHER INFORMATION: megAI, KS3
NAME/KEY: misc_feature
LOCATION: (42544)...(42581)
OTHER INFORMATION: megAI, AT3
NAME/KEY: misc_feature
LOCATION: (42630)...(42673)
OTHER INFORMATION: megAI, KR3 (inactive)
NAME/KEY: misc_feature
LOCATION: (42698)...(42728)
OTHER INFORMATION: megAI, ACP3
NAME/KEY: misc_feature
LOCATION: (42733)...(42850)
OTHER INFORMATION: megAI, KS4
NAME/KEY: misc_feature
LOCATION: (42897)...(42931)
OTHER INFORMATION: megAI, AT4
NAME/KEY: misc_feature
LOCATION: (42953)...(43047)
OTHER INFORMATION: megAI, DH4
NAME/KEY: misc_feature
LOCATION: (43136)...(43244)
OTHER INFORMATION: megAI, BR4
NAME/KEY: misc_feature
LOCATION: (43257)...(43279)
OTHER INFORMATION: megAI, KR4
NAME/KEY: misc_feature
LOCATION: (43052)...(43312)
OTHER INFORMATION: megAI, ACP4
NAME/KEY: CDS
LOCATION: (43666)...(43271)
OTHER INFORMATION: megAI, SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc_feature
LOCATION: (43780)...(435027)
OTHER INFORMATION: megAI, KS5
NAME/KEY: misc_feature
LOCATION: (43585)...(436419)
OTHER INFORMATION: megAI, AT5
NAME/KEY: misc_feature
LOCATION: (437068)...(437604)
OTHER INFORMATION: megAI, KR5
NAME/KEY: misc_feature
LOCATION: (437850)...(438120)
OTHER INFORMATION: megAI, ACP5
NAME/KEY: misc_feature
LOCATION: (438187)...(439470)
OTHER INFORMATION: megAI, KS6
NAME/KEY: misc_feature
LOCATION: (439755)...(440811)
OTHER INFORMATION: megAI, AT6
NAME/KEY: misc_feature
LOCATION: (44106)...(441936)
OTHER INFORMATION: megAI, KR6
NAME/KEY: misc_feature
LOCATION: (442168)...(442425)
OTHER INFORMATION: megAI, ACP6
NAME/KEY: misc_feature
LOCATION: (44285)...(443271)
OTHER INFORMATION: megAI, TE
```











;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 13605  
;; LENGTH: 804  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13605

Query Match  
Best Local Similarity 47.2%; Score 38.8; DB 4; Length 804;  
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1043 CCGGACGGGCTGCGCTGAGATCTGCTCAGAGACGTGACCCGCTGTGAGGG 1102  
DB 303 CCGGCTGGGGCGGACAGCATCTGCGCTGCGCACTTCCGCGCGGATGCGGCC 362  
QY 1103 GGTGCTGGCCCCAGCCCCCGCGCTGGCCCTGGGCTGCCAAGATGACATGCCAG 1162  
DB 363 AGGTGGGACATGAGCGGATGACACGCGCGCGGATTCGCGTCTATCCGTGGCCAG 422  
QY 1163 TCACGAGCCCGCTCTGTCACAGGATGACCTGAGTCCCTTCTGAGGAGACACTTCA 1222  
DB 423 TCCAAAGGCGGCACTTCCGCTGTCATCAGTGTGCTGCTTTGAGCAGGCCCTCG 482  
QY 1223 TGGCATCTGCAAGTGGGCAATCCAGAGATGCGCCGCGCGGCCCTTCCCTCTG 1282  
DB 483 CCGCGAGCGGACGCGCGCGGAGAGCGCGCTGCGCGGTTGAGAAAGCGGCTCGGCGA 542  
QY 1283 ACCCCAGATG 1292  
DB 543 GCGCTGCTG 552

RESULT 56  
US-09-252-991A-13494  
; Sequence 13494, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13494  
; LENGTH: 984  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13494

Query Match  
Best Local Similarity 47.2%; Score 38.8; DB 4; Length 984;  
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1043 CCGGACGGGCTGCGCTGAGATCTGCTCAGAGACGTGACCCGCTGTGAGGG 1102  
DB 145 CCGGCTGGGGCGGACAGCATCTGCGCTGCGCACTTCCGCGCGGATGCGGCC 204  
QY 1103 GGTGCTGGCCCCAGCCCCCGCGCTGGCCCTGGGCTGCCAAGATGACACTCCAG 1162  
DB 205 AGGTGGGACATGAGCGGATGACACGCGCGCGGATTCGCGTCTATCCGTGGCCAG 264  
QY 1163 TCACGAGCCCGCTCTGTCACAGGATGACCTGAGTCCCTTCTGAGGAGACACTTCA 1222  
DB 265 TCCAAAGGCGGCACTTCCGCTGTCATCAGTGTGCTGCTTTGAGCAGGCCCTCG 324

QY 1223 TGGCATCTGCAAGTGGGCAATCCAGAGATGCGCCGCTCCGCGGCCCTTCCCTCTG 1282  
DB 325 CCGGAGCGGACGCGGCGGAGCGCGCTGCGCGGTTGAGAAAGCGGCTCTCGGCA 384  
QY 1283 ACCCCAGATG 1292  
DB 385 GCGCTGCTG 394

RESULT 57  
US-09-252-991A-4137/c  
; Sequence 4137, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4137  
; LENGTH: 1104  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4137

Query Match  
Best Local Similarity 44.2%; Score 38.8; DB 4; Length 1104;  
Matches 160; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 360 GGTCAAGGAGAAATTCAGGCGCTGCGAGCATGAGGAGGAGGAGAGACTGAGGAGTTCA 419  
DB 522 GACATCGGAGATCATCTCTGCTCGCGGCGCGCTGAGATGATGACATGATGATG 463  
QY 420 GGTACCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479  
DB 462 AGGTGCGGAGAGCGCGCGCGGAGAAATTCGCGCTGCGATGCGCTGCGCGCGGCGG 403  
QY 480 CTGCAACCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539  
DB 402 GACATCGGAGAACTTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 343  
QY 540 TGTGCGGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 599  
DB 342 GCGGAGGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 283  
QY 600 CACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659  
DB 282 TCCCTGAGCTTACGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 223  
QY 660 AGCTGCTGAGGAGCAACTCCAGAGGTGACGCCCGGAGAGAGAGGAGGAGGAGG 719  
DB 222 TTGCGGCTTATCTCGGCGGTCTGCGGCGGTCTGCGGCGGTCTGCGGCGGTCTG 163  
QY 720 GA 721  
DB 162 GA 161

RESULT 58  
US-09-252-991A-4071  
; Sequence 4071, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136



APPLICANT: Huang, Huang  
TITLE OF INVENTION: Generation of Human Cytomegalovirus Yeast Artificial Chromosome  
FILE REFERENCE: 98,299  
CURRENT APPLICATION NUMBER: US/09/765,400  
CURRENT FILING DATE: 2000-11-03  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 64  
LENGTH: 229354  
TYPE: DNA  
ORGANISM: Human cytomegalovirus;  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Human cytomegalovirus strain AD169 (GenBank X17403.1)  
US-09-765-400-64

Query Match 2.5%; Score 38.6; DB 4; Length 229354;  
Best Local Similarity 49.8%; Pred. No. 21;  
Matches 124; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

Qy 946 GCGGGGTGTGCGAGATGTGATCGAGCTGTGCTGTGTACTCACTGCGCTGCTGCTTCC 1005  
Db 93012 GCGAGATCGCGGCGAGGTTCCGGCGGTGCTGCTGCGAGCTGCGCGCTCCCGCC 93071  
Qy 1006 ACTGGGCTGCCACTTCCAGCCGCACTCCCGCCCGGAGCGGCGCTGCGTCAAT 1065  
Db 93072 TGCCTCCCGCTTACCCCACTCCCGCGGAGCGGCGCGCTGCTCCCGG 93131  
Qy 1066 CCGTCTCAGAGAGCTGATCCCGCCCTGTGTGAGGGGGTGTGCGCCCGAGCCCGCC 1125  
Db 93132 CCGTCTCCACCGCCCTG-GAGCACATCCGGGCGCTGTGCGCGGAGCGGCGCGCC 93190  
Qy 1126 GCCTGGCCCTGTGCGCTGCGCAAGATGACATCGACGAGCCCGCTGTGACAGG 1185  
Db 93191 GCTCGGAGCTCGCGCGGGGCTCCCTCCCTCCCGCGTCAACCCCGATCCAGCGG 93250  
Qy 1186 ATGACCTGG 1194  
Db 93251 CCGGCGGG 93259

RESULT 62  
US-09-705-400-64  
Sequence 64, Application US/09705400  
Patent No. 6692954  
GENERAL INFORMATION:  
APPLICANT: Ghazal, Peter  
APPLICANT: Huang, Huang  
TITLE OF INVENTION: Generation of Human Cytomegalovirus Yeast Artificial Chromosome  
FILE REFERENCE: 98,299  
CURRENT APPLICATION NUMBER: US/09/705,400  
CURRENT FILING DATE: 2000-11-03  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 64  
LENGTH: 229354  
TYPE: DNA  
ORGANISM: Human cytomegalovirus  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Human cytomegalovirus strain AD169 (GenBank X17403.1)  
US-09-705-400-64

Query Match 2.5%; Score 38.6; DB 4; Length 229354;  
Best Local Similarity 49.8%; Pred. No. 21;  
Matches 124; Conservative 0; Mismatches 124; Indels 1; Gaps 1;  
Qy 946 GCGGGGTGTGCGAGATGTGATCGAGCTGTGCTGTGTACTCACTGCGCTGCTTCC 1005  
Db 93012 GCGAGATCGCGGCGAGGTTCCGGCGGTGCTGCTGCGAGCTGCGCGCTCCCGCC 93071

Qy 1006 ACTGGGCTGCCACTTCCAGCCGCACTCCCGCGGAGCGGCGCTGCGTCAAT 1065  
Db 93072 TGCCTCCCGCTTACCCCACTCCCGCGGAGCGGCGCGCTGCTCCCGCC 93131  
Qy 1066 CCGTCTCAGAGAGCTGATCCCGCCCTGTGTGAGGGGGTGTGCGCCCGAGCCCGCC 1125  
Db 93132 CCGTCTCCACCGCCCTG-GAGCACATCCGGGCGCTGTGAGCGGAGCGGCGCGCC 93190  
Qy 1126 GCCTGGCCCTGTGCGCTGCGCAAGATGACATCGACGAGTCCCGCTGTGACAGG 1185  
Db 93191 GCTCGGAGCTCGCGCGGGGCTCCCTCCCTCCCGCGTCAACCCCGATCCAGCGG 93250  
Qy 1186 ATGACCTGG 1194  
Db 93251 CCGGCGGG 93259

RESULT 63  
US-09-724-797-21  
Sequence 21, Application US/09724797  
Patent No. 673398  
GENERAL INFORMATION:  
APPLICANT: Jon S. THORSON  
TITLE OF INVENTION: MICROMOSPORE ECHINOSPORE GENES  
TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF  
FILE REFERENCE: 2653-40  
CURRENT APPLICATION NUMBER: US/09/724,797  
CURRENT FILING DATE: 2000-11-28  
PRIOR FILING DATE: 1998-12-07  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 1683  
TYPE: DNA  
ORGANISM: Bacteria  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1683)  
US-09-724-797-21

Query Match 2.5%; Score 38.4; DB 4; Length 1683;  
Best Local Similarity 48.6%; Pred. No. 4.5;  
Matches 105; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
Qy 555 GACGGGGGAGCTATCTGTGACGAGCTGCTCGGCGCTTCACCTGAGCTGCTG 614  
Db 41 GCTGCGGGGACCTGCGGCTGCTGCGGCGCGCCCTCAACAGGCGGACCGCGG 100  
Qy 615 TCCCTCGCTCGGAGATCCCAAGTGGAGCTGAGAGTGTCCAGCTGCGAGGA 674  
Db 101 CCGCTGTGCGCGCGGCGGAGCGCAGTGTGCTGCGGTGACCGGCGAGTATCAGCT 160  
Qy 675 ACAATCCAGAGTCAAGCCCGGCGGAGAGAGCCCGGCGGAGAGCCAGAGCCG 734  
Db 161 ACCGCGGAGACCGACCGGCGGAGTGGCGGCTGTGCGGCTGTGCGAGCTG 220  
Qy 735 ACCCGCTCGCGGCGGCTGTGAGTGGCGGAGAG 770  
Db 221 CCGAGTCCCTCGACCTACACAGCGCGGCTGCGG 256

RESULT 64  
US-09-077-940A-3  
Sequence 3, Application US/09077940A  
Patent No. 6576441  
GENERAL INFORMATION:  
APPLICANT: KIMURA, Toru et al.  
TITLE OF INVENTION: NOVEL SEMAPHORIN 2 AND GENE ENCODING THE SAME  
FILE REFERENCE: 0020-4426P  
CURRENT APPLICATION NUMBER: US/09/077,940A  
CURRENT FILING DATE: 1998-06-05



NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 3524  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(38)  
OTHER INFORMATION:  
NAME/KEY: 3'UTR  
LOCATION: (2706)..(3524)  
OTHER INFORMATION:  
NAME/KEY: CDS  
LOCATION: (39)..(2702)  
OTHER INFORMATION:  
US-09-077-940A-3

Query Match 2.5%; Score 38.4; DB 4; Length 3524;  
Best Local Similarity 47.0%; Pred. No. 5.8;  
Matches 154; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 423 ACCGAGATGCTGCTGGGGAGCTGTTTGGAGAGAGTGCTCTCAGAGGGTCTG 482  
DB 1887 ACGTGGGCTGGTTCGTGGGCTCCCTGAGCGCGGAGCTGCGCGGCAAGGACAG 1946  
QY 483 CACCCAGCC---CAGTGTGATGGGCGTCTTGGCCAGAAAGATAGAGCGAG 539  
DB 1947 GAGGCGATCTGGCGGAGCGGGGCGAGCGGCTGAGCTCAGCCGCTGGGCGAG 2006  
QY 540 TGTGCGTGTGTCGGGACGGCGGGAGCTCATCTGTGAGCGGCTGCGGCGCTTC 599  
DB 2007 CCGAGGGCGCAGAGGTCCCGGGGGCGGGGCGGAGGCGGTGGCGCGCGGGTTC 2066  
QY 600 CACCTGGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAAGTGGAGCTGAGGTCTCC 659  
DB 2067 CCGAGGCTCTGCTGCGCGCCCTGATGAGAAAGGCTGGGCGCAAGGCGCTGCTGAG 2126  
QY 660 AGCTGCTGCGAGCAAGTCCAGAGAGTGACGCCCGGGCAGAGAGCGCGGCCAG 719  
DB 2127 GCGGGGCGCCAGCAGCTGAGCTGGGGCTGCTGCCACCGCGAGAGCGCGCTGCCG 2186  
QY 720 GAGCCACCCGCTGAGAGCCCGGCTCCCGC 747  
DB 2187 CAGAGCGCTGCGCCGCTCCGAGCCGCGC 2214

## RESULT 65

US-10-140-002-543  
Sequence 543, Application US/10140002  
Patent No. 6725730  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeFoige, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvarcoff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Guirney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3350R1C59  
CURRENT APPLICATION NUMBER: US/10/140,002

CURRENT FILING DATE: 2002-05-06  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 543  
LENGTH: 3721  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-140-002-543

Query Match 2.5%; Score 38.4; DB 4; Length 3721;  
Best Local Similarity 47.0%; Pred. No. 5.9;  
Matches 154; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 423 ACCGAGATGCTGCTGGGGAGCTGTTTGGAGAGAGTGCTCTCAGAGGGTCTG 482  
DB 1887 ACGTGGGCTGGTTCGTGGGCTCCCTGAGCGCGGAGCTGCGCGGCAAGGACAG 1926  
QY 483 CACCCAGCC---CAGTGTGATGGGCGTCTTGGCCAGAAAGATAGAGCGAG 539  
DB 1927 GAGGCGATCTGGCGCAGCGGGCGGAGCGGCTGAGCGTCAAGCGCTGCGGCGAG 1986  
QY 540 TGTGCGTGTGTCGGGACGGCGGGAGCTCATCTGTGAGCGGCTGCGGCGCTTC 599  
DB 1987 CCGAGGGCGCAGAGGTCCCGGGGGCGGGGCGGAGGCGGTGGCGCGCGGGTTC 2046  
QY 600 CACCTGGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAAGTGGAGCTGAGGTCTCC 659  
DB 2047 CCGAGGCGCTGCTGCTGCGCCCTGATGAGAAAGGCTGGGCGCAAGCGCTGCGAG 2106  
QY 660 AGCTGCTGCGAGCAAGTCCAGAGAGTGACGCCCGGGCAGAGAGCGCGGCCAG 719  
DB 2107 GCGGGGCGCCAGCAGCTGGAATCGGGGCTGCTGCGCCAGCGCGAGCAGCGCGCTGCCG 2166  
QY 720 GAGCCACCCGCTGAGAGCCCGGCTCCCGC 747  
DB 2167 CAGAGCGCTGCGCCGCTCCGAGCCGCGC 2194

## RESULT 66

US-09-191-171-7/C  
Sequence 7, Application US/09191171  
Patent No. 6149909  
Patent No. 6149909 6143294  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/191,171  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995

APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU92/00611  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: DIGIGLIO, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4480 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-191-171-7

Query Match  
Best Local Similarity 47.2%; Pred. No. 6.3;  
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

2.5%; Score 38.4; DB 3; Length 4480;  
Best Local Similarity 47.2%; Pred. No. 6.3;  
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

592 GGGCCCTTCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAATGGAGACTGGA 651  
Db 2618 GGAACGCCACCGGTGGTGTCTGTCCAGACGGTCCCGCTGCGACACTTCGCGCCAGA 2559  
QY 652 GGTGCTCAGCTGCTGCTGAGGCAAGTCCAGAGGTGAGCCCGGAGAGAGAGAGAGAG 711  
Db 2558 GCTGCTCTCAT--CTGG 2501  
QY 712 GGGCCCGAG 771  
Db 2500 CGCCCCCGAG 2441  
QY 772 AGTAAGAGGTCCACCTGGGGAACCCCTAGCGGAGATGAGACACTTGTCTACAAGC 831  
Db 2440 TCCCGAG 2381  
QY 832 ACCTGCGGCTCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 2380 TCCCGGAG 2321  
QY 892 CCTACTGTGTGGGTC 909  
Db 2320 CCCGAGGCGCGCGGCTC 2303

RESULT 67  
US-09-385-707-7/c  
Sequence 7, Application US/09385707  
GENERAL INFORMATION:  
PATENT NO. 6238662  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOBWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDIIONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City

STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,707  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/494,104  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: DIGIGLIO, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4480 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-707-7

Query Match  
Best Local Similarity 47.2%; Pred. No. 6.3;  
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

2.5%; Score 38.4; DB 3; Length 4480;  
Best Local Similarity 47.2%; Pred. No. 6.3;  
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

592 GGGCCCTTCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAATGGAGACTGGA 651  
Db 2618 GGAACGCCACCGGTGGTGTCTGTCCAGACGGTCCCGCTGCGACACTTCGCGCCAGA 2559  
QY 652 GGTGCTCAGCTGCTGCTGAGGCAAGTCCAGAGGTGAGCCCGGAGAGAGAGAGAGAG 711  
Db 2558 GCTGCTCTCAT--CTGG 2501  
QY 712 GGGCCCGAG 771  
Db 2500 CGCCCCCGAG 2441  
QY 772 AGTAAGAGGTCCACCTGGGGAACCCCTAGCGGAGATGAGACACTTGTCTACAAGC 831  
Db 2440 TCCCGAG 2381  
QY 832 ACCTGCGGCTCGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 2380 TCCCGGAG 2321  
QY 892 CCTACTGTGTGGGTC 909  
Db 2320 CCCGAGGCGCGCGGCTC 2303

RESULT 68  
US-09-639-696C-6/c  
Sequence 6, Application US/09639696C  
PATENT NO. 6524835  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOBWOOD, John J.



Best Local Similarity 46.7%; Pred. No. 5.5; Matches 121; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 974 GCTGCGGTACTCACTGCGCCCTCTTCCATGCGCGTGCACCTTCCACCGGAC 1033  
Db 847 GCTGCGGTACTCACTGCGCCCTCTTCCATGCGCGTGCACCTTCCACCGGAC 966  
QY 1034 CTCGCGCGCGCGCGCGCGCTGCTGCGTGCAGATCTTCTGAGAGACGTGACCCGACCC 1093  
Db 907 CGTCCCGCGCGCGCGCGCTGCTGCGTGCAGATCTTCTGAGAGACGTGACCCGACCC 966  
QY 1094 TGTGAGAGGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153  
Db 967 GTTCTCTTCCCG 1026  
QY 1154 CACTGCGAGTCAAGACCGCGCTGCGTGCAGAGATGACTGAGTCTTCTGAGCGAC 1213  
Db 1027 CTATCCCG 1086  
QY 1214 CACCTTGCATGGCATCTG 1232  
Db 1087 GCTCTTGGCTCTGCGCTG 1105

RESULT 71  
US-08-289-112-1  
; Sequence 1, Application US/08289112  
; Patent No. 5688640  
; GENERAL INFORMATION:  
; APPLICANT: Yamagisawa, Masashi  
; TITLE OF INVENTION: Endothelin Converting Enzyme-1. A  
; TITLE OF INVENTION: Membrane-Bound Metalloprotease That Catalyzes The  
; NUMBER OF INVENTION: Proteolytic Activation of Big Endothelin-1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/289,112  
; FILING DATE: 10-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSD:414/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2889 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 118..2391  
; US-08-289-112-1

Query Match 2.5%; Score 38.2; DB 1; Length 2889;  
Best Local Similarity 51.5%; Pred. No. 6.1;  
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 567 CTCATCTGTGTGACGCGTCCCTCGGCGCTTCCACTGCGCTGCTGCTCCGCTC 626  
Db 157 CTCCTGGGAGAGAGGCGCCCTGCGGCTCAAGTCTCTCCGCTTCCGCTCTTCCCTG 216  
QY 627 CCGGAGATCCCAATGAGACCTGAGAGTCTTCCAGTGTCTGCGGCAACAGTCCAGAG 686  
Db 217 CAGGTAACTTCCGAGGCGCGCGGAGCGGCAAGATGCTGAGCGCGCGGAGAGCCGCGT 276  
QY 687 GTGACGCCCGGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737  
Db 277 GAGAACCGGCTGT 327

RESULT 72  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, c, g or g  
; US-09-103-840A-2

Query Match 2.5%; Score 38.2; DB 3; Length 4403765;  
Best Local Similarity 44.8%; Pred. No. 58;  
Matches 187; Conservative 0; Mismatches 228; Indels 2; Gaps 1;

QY 681 CAGAGGTGACG 740  
Db 3936918 CCG 3936859  
QY 741 CTCCTCCCG 800  
Db 3936858 GTGCG 3936799  
QY 801 GCCGCGATGACAGACGCTTGTCTTACAGACGCTGCGCGCGCGCGCGCGCGCGCG 860  
Db 3936798 TGACCAACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936739  
QY 861 GTGCGAGTGTGAGTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920  
Db 3936738 CCG 3936679  
QY 921 CAGAACCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 980  
Db 3936678 GTGCG 3936619  
QY 981 TGTACTCA--CTGCG 1038  
Db 3936618 GTGCG 3936559  
QY 1039 GCGCGCGAGCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1095  
Db 3936558 CCGCGAGCG 3936502

```
RESULT 73
US-08-458-568A-3/C
; Sequence 3, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3439
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..702
; US-08-458-568A-3

Query Match          2.5%; Score 38; DB 1; Length 702;
Best Local Similarity 50.5%; Pred. No. 4.2;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 989 CTGCGCGCTGCTTCACCTGCGCTGCGCACTTCCAGCGCGACCTCCCGCGCGGAGC 1048
DB 400 CAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 341
QY 1049 GGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
DB 340 CGGGGGGGGGGGTGAACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 281
QY 1109 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1168
DB 280 TCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 221
QY 1169 GC 1170
DB 220 GC 219
```

```
RESULT 74
PCT-US91-06532-1
; Sequence 1, Application PC/TUS9106532
; GENERAL INFORMATION:
; APPLICANT: Rolzman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; TITLE OF INVENTION: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marehall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 19910910
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US91-06532-1

Query Match          2.5%; Score 38; DB 5; Length 1335;
Best Local Similarity 50.5%; Pred. No. 5.3;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 989 CTGCGCGCTGCTTCACCTGCGCTGCGCACTTCCAGCGCGACCTCCCGCGCGGAGC 1048
DB 407 CAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 466
QY 1049 GGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
DB 467 CGGGGGGGGGGGCTAACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 526
QY 1109 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1168
DB 527 TCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 586
QY 1169 GC 1170
DB 587 GC 588

RESULT 75
US-08-196-218-31
; Sequence 31, Application US/08196218
; Patent No. 5614619
; GENERAL INFORMATION:
; APPLICANT: Piepersberg, Wolfgang
; APPLICANT: Stockmann, Michael
; APPLICANT: Taleghani, Kampanz Mansouri
```

APPLICANT: Dietler, Jurgen  
APPLICANT: Grabley, Susanne  
APPLICANT: Sichel, Petra  
APPLICANT: Brau, Barbara  
TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes  
TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/196,218  
FILING DATE: 25-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ogden, Scasia L.  
REGISTRATION NUMBER: 36,228  
REFERENCE/DOCKET NUMBER: 02481.1372-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2634 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..401  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 416..1531  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1561..2625  
US-08-196-218-31  
Query Match  
Best Local Similarity 51.8%; Score 38; DB 1; Length 2634;  
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 968 GGACGTGCTGGGATGATCACTGCGCGCCCTTCCATCTGCGCTGCACCTTCCAC 1027  
DB 2028 GGAAGAGCCGTGACAGCCGCGAGACATGCGCTGATCGGGTCTTCTTACCCGC 2087  
QY 1028 CGGCACTTCCCGGCGGAGCGGCGCTGCGCTGCAGATCTCTGAGAGAGTGAACCC 1087  
DB 2088 CGCCATTCACCAAGGCGGTGGCGCCATCTCGCCAGACCGCGGGAATGGAATCAG 2147  
QY 1088 AGCCCTGTGAGAGGGGTGCTGCGCCCAAGCCCGCCGCTGAGCC 1133  
DB 2148 CGACGCGCTTCAGTGTGTGATCACTCGGCGGAGAGTGGCGCC 2193  
RESULT 76  
US-08-681-953-31  
Sequence 31, Application US/08681953  
Patent No. 5710032  
GENERAL INFORMATION:  
APPLICANT: Piepersberg, Wolfgang

APPLICANT: Stockmann, Michael  
APPLICANT: Taleghani, Kamiz Mansouri  
APPLICANT: Dietler, Jurgen  
APPLICANT: Grabley, Susanne  
APPLICANT: Sichel, Petra  
APPLICANT: Brau, Barbara  
TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes  
TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/681,953  
FILING DATE: 30-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/196,218  
FILING DATE: 25-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ogden, Scasia L.  
REGISTRATION NUMBER: 36,228  
REFERENCE/DOCKET NUMBER: 02481.1372-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2634 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..401  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 416..1531  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1561..2625  
US-08-681-953-31  
Query Match  
Best Local Similarity 51.8%; Score 38; DB 1; Length 2634;  
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 968 GGACGTGCTGGGATGATCACTGCGCGCCCTTCCATCTGCGCTGCACCTTCCAC 1027  
DB 2028 GGAAGAGCCGTGACAGCCGCGAGACATGCGCTGATCGGGTCTTCTTACCCGC 2087  
QY 1028 CGGCACTTCCCGGCGGAGCGGCGCTGCGCTGCAGATCTCTGAGAGAGTGAACCC 1087  
DB 2088 CGCCATTCACCAAGGCGGTGGCGCCATCTCGCCAGACCGCGGGAATGGAATCAG 2147  
QY 1088 AGCCCTGTGAGAGGGGTGCTGCGCCCAAGCCCGCCGCTGAGCC 1133  
DB 2148 CGACGCGCTTCAGTGTGTGATCACTCGGCGGAGAGTGGCGCC 2193  
RESULT 77





```
QY 500 CATGGGCGTCTCTTGGAGAGAGTGGCTGCGGAGAGTGGCTGCTGCTGGAGCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9951 GGAGGACGAGATCCGGCTGTGCTGGAGAGCGCGGTACGCGGACGTCTGGACGT 10010
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTGGGGCTTCCACTGGGCC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 10011 CGCCCGCTCAACGAGCGCCGAGCGCCGCTCTGTCTCGGCGAGCGCGAC 10059

RESULT 80
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTGGAGAGAGTGGCTGTCAGAGAGGCTGCTGACCCAGCCACTG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11632 GCGCGCGCGCTCATGACGAGAGTCCGCCGCTGGCGAGTCTGCGCTCCAGCCGC 11691
QY 500 CATGGCGCTCTTGGCTGCTGCCAAGAGATGAGAGCGAGTGTGCTGTGCGGACGG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11692 GGAGGACGAGATCCGGCTGTGCTGGAGAGCGAGGACGGTACGCTGAGACGT 11751
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTGGGGCTTCCACTGGGCC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11752 CGCCCGCTCAACGAGCGCCGAGCGCCGCTCTGTCTCGGCGAGCGCGAC 11800

RESULT 81
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
```

```
EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTGGAGAGAGTGGCTGTCAGAGAGGCTGCTGACCCAGCCACTG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9774 GCGCGCGCGCTCATGACGAGAGTCCGCCGCTGGCGAGTCTGCGCTCCAGCGCC 9833
QY 500 CATGGCGCTCTTGGCTGCTGCCAAGAGATGAGAGCGAGTGTGCTGTGCGGACGG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9834 GGAGGACGAGATCCGGCTGTGCTGGAGAGCGAGGACGGTACGCGGACGTCTGACGT 9893
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTGGGGCTTCCACTGGGCC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9894 CGCCCGCTCAACGAGCGCCGAGCGCCGCTCTGTCTCGGCGAGCGCGAC 9942

RESULT 82
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
```

RESULT 83  
US-09-657-440-19  
Sequence 19, Application US/09657440  
Patent No. 6509455  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Melanie C.  
APPLICANT: BETLACH, Mary C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002120  
CURRENT APPLICATION NUMBER: US/09/657,440  
CURRENT FILING DATE: 2000-09-07  
PRIOR APPLICATION NUMBER: 09/320,878  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 38506  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 2.4%; Score 37.8; DB 4; Length 38506;  
Best Local Similarity 51.5%; Pred. No. 18;  
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 440 GGGGAGCTGTTTGGAGAGAGGTGAGCTCTCAGAGAGGTGTCGACCCAGCCAGCTCTG 499  
DB 9774 GGGCGCCCGGCTATGCAAGAGCTGCCCCGGTGGCGGATGCTGCGCTCCAGGCCCG 9833  
QY 500 CATGGCGTCTCTTGGCTTGTGCGAGAGATGAGACGAGTGTGCGGTGTGCGGAGCG 559  
DB 9834 GAGGAGCGAGATCCGCTGTGCGAGACGAGAGCCGTAACGCGGAGCGTCTGAGCGT 9893  
QY 560 CGGGAGCTCATCTGTGTGAGCGGCTCCCTCGGGCTTCCACTGCGC 608  
DB 9894 CGCGCGCTCAACGCGCCGAGCGCGCTCTGTCTCGGAGCGCGAC 9942

RESULT 84  
US-09-305-640-1/C  
Sequence 1, Application US/09305640B  
Patent No. 6255468  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham plc  
TITLE OF INVENTION: No. 6255468el Compounds  
FILE REFERENCE: GP30124  
CURRENT APPLICATION NUMBER: US/09/305,640B  
CURRENT FILING DATE: 1999-05-05  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2560  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-305-640-1

Query Match 2.4%; Score 37.6; DB 3; Length 2560;  
Best Local Similarity 46.8%; Pred. No. 83; Indels 0; Gaps 0;  
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 397 CAGGCGAGAGACTGGGAGTTCAAGTACCCAGAGATCTGCTGGGAGCTGTTTGGGA 456  
DB 514 CGGAGAGGCG 455  
QY 457 AGGAGTGGCTCTCAGAGAGGTGCTGACCCCGAGCTGTCATGAGGCGTCTTGGC 516  
DB 454 CCAGATGGCGGAGAGCG 395

QY 517 TGTGCCAGAAAGATGAGACGAGTGTGCCGTGTGTGCGAGACGCGGGAGCTCATCTGCT 576  
DB 394 GGTTCAGCGCGGAGCGCCCGAGCCGGGCCCCGGTGGCTGTGCGCCAGCGCCCAACGGGA 335  
QY 577 GTGACGGCTGCGCTTGGGCGCTTTCACCTGAGCTGCTGCTCCCTCGCGGAGATCC 636  
DB 334 AGCCCGGGGAGAGAGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 275  
QY 637 CCAGTGGGACCT 648  
DB 274 CCTTTGGAAT 263

RESULT 85  
US-10-140-002-119/C  
Sequence 119, Application US/10140002  
Patent No. 6725730  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Matanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zeman  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3330R1C59  
CURRENT APPLICATION NUMBER: US/10/140,002  
CURRENT FILING DATE: 2002-05-06  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 119  
LENGTH: 2868  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-140-002-119

Query Match 2.4%; Score 37.6; DB 4; Length 2868;  
Best Local Similarity 46.8%; Pred. No. 86; Indels 0; Gaps 0;  
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;  
QY 397 CAGGCGAGAGACTGGGAGTTCAAGTACCCAGAGATGCTGCTGGGAGACTGTTTGGGA 456  
DB 494 CGGAGAGGCG 435  
QY 457 AGGAGTGGCTCTCAGAGAGGTGCTGACCCCGAGCGAGTGTGAGGCGCGACATAGCCG 516  
DB 434 CCAGATGGCGGAGAGCG 375  
QY 517 TGTGCCAGAAAGATGAGACGAGTGTGCCGTGTGTGCGAGACGCGGGAGCTCATCTGCT 576  
DB 374 GGTTCAGCGCGGAGCGCCCGAGCCGGGCCCCGGTGGCTGTGCGCCAGCGCCCAACGGGA 315  
QY 577 GTGACGGCTGCGCTTGGGCGCTTTCACCTGAGCTGCTGCTCCCTCGCGGAGATCC 636  
DB 314 AGCCCGGGGAGAGAGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 255  
QY 637 CCAGTGGGACCT 648  
DB 254 CCTTTGGAAT 243

RESULT 86  
US-09-774-528-383/c  
Sequence 383, Application US/09774528  
Patent No. 6743619  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Zhou, Ping  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyan  
APPLICANT: Zhang, Jie  
APPLICANT: Zhao, Qing A.  
APPLICANT: Yang, Yonghong  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wang, Dunhui  
APPLICANT: Dimauc, Radoje T.  
TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
FILE REFERENCE: 802  
CURRENT APPLICATION NUMBER: US/09/774,528  
CURRENT FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 441  
SOFTWARE: pc\_fl\_genes Version 2.0  
SEQ ID NO 383  
LENGTH: 1706  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (92)..(1309)  
US-09-774-528-383

Query Match 2.4%; Score 37.4; DB 4; Length 1706;  
Best Local Similarity 50.9%; Pred. No. 8.1;  
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 627 CGGAGATCCCGAGGAGCTGAGGCTCCAGCTGCTCAGGCAACAGTCCAGAG 686  
DB 1199 CGCGAGGCCCGAGAGGGTGAATGGGGTCTTCGAGGGGGCTGGGTCATCATGAAAGTG 1140  
QY 687 GTGCAAGCCCGGAGAGAGAGCCCGGCGCCCAAGAGCCCGGTGAGACCCCGCTCCC 746  
DB 1139 CTGAAGCTTCCTGAGAGAGGTGCCAGGAGAGAGGCCACCTGACACTATCTGGGGCCC 1080  
QY 747 CGGGGGCTTAGGTCGGCGGAGAGAGGTGAAGGTCCAGCTGGGGAAACCCCTAG 801  
DB 1079 CAGAGGCTGAGACGAGCGCTCTGAGTCAAGAGAGGTGGCTGAGAGGCCCTGG 1025

RESULT 87  
US-09-252-991A-9715/c  
Sequence 9715, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 9715  
LENGTH: 648  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9715

Query Match 2.4%; Score 37.2; DB 4; Length 648;  
Best Local Similarity 46.5%; Pred. No. 6.5;  
Matches 120; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 409 TGGGAGTTCAAGTACCCAGAGATGCTGCGGGAGAGCTGTTTGGAAAGAGGTGGCTC 468  
DB 463 TGGGCTCAGCGGCTTGGCGGATGCGCTGGGTGAGGCGGAGAGAGGTCT 404  
QY 469 TCAGAGAGGTGTCGACCCAGCCAGCTGATGAGGAGTCTTTCCTGTGCCAAGA 528  
DB 403 GTCCAGGCGCATGATGACTGACGCGGCTTCAAGTCTCAGCGGCTGGCTGCGA 344  
QY 529 ATGAGACGAGTGTGCTGTGTGCGGACGCGGAGAGCTCATCTGTGTGACGCTGCC 588  
DB 343 TGTAGCCGAGGCGCGCTGCGCAGAGACCGCTTGGCAGACGCGCGCTGAACACCG 284  
QY 589 CTCGGGCTTCCACCTGCGGCTGCTCCCTCCGCTCCGAGATCCCGAGTGGAGCT 648  
DB 283 AGTAGACCATCACTTCAATTGCGGGCAGTCCGGGTGAGCCATTTCACAGGTGATAC 224  
QY 649 GAGGTGCTCCAGCTGCC 666  
DB 223 CCTCGCGCTGCGCTGCC 206

RESULT 88  
US-08-318-837-1  
Sequence 1, Application US/08318837  
Patent No. 5981277  
GENERAL INFORMATION:  
APPLICANT: FRANKEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,  
APPLICANT: ANDRE; VAN HEUVERSWYN, HUGO  
TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID  
TITLE OF INVENTION: COOKING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERAPY  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESS: BIERMAN AND MUSERLIAN  
STREET: 600 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,837  
FILING DATE: 13-OCT-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP 93/01022  
FILING DATE: 28-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 92,401,231.3  
FILING DATE: 30-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CHARLES A. MUSERLIAN  
REGISTRATION NUMBER: 19,683  
REFERENCE/DOCKET NUMBER: 410,007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 661-8000  
TELEFAX: (212) 661-8002  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2150 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

```
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Spleen tissue from healthy human
FEATURE:
NAME/KEY: CDS
LOCATION: 1810..1982
FEATURE:
NAME/KEY: misc_feature
LOCATION: 543
OTHER INFORMATION: C may be present or absent
FEATURE:
NAME/KEY: Intron
LOCATION: 1980..2150
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..1809
US-08-318-837-1
```

Query Match 2.4%; Score 37.2; DB 2; Length 2150;  
Best Local Similarity 51.2%; Pred. No. 9.8;  
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
QY 584 CTGCGCTGGGCTTCCACCTGGCTGCTCCCTCCGCTCGGAGATCCCGAGTGG 643
DB 1709 CCGAGCTCAGAGCCGCCGCCGCCGCCGCCGCTCCCGAGTCTGCTCCGGGG 1768
QY 644 GACCTGAGAGTGTCTCCAGCTGCTGAGGCAACATGTCAGAGAGTGCAGCCCGGGCAGA 703
DB 1769 GTTCGCGAGCGCGGGGCGCGGCGGAGCGCGGCGGAGCATGCGGGCGCGCGGGG 1828
QY 704 GGAGCCCCCGCCCGAGAGCCCGTGAAGACCCCGCTCCCGCGGGGC 753
DB 1829 CGGCTGGGAGCGCGGCGGAGCGCTGTGCGCGGAGCCCGCGCGGGC 1878
```

RESULT 89  
US-09-428-711A-13/C  
Sequence 13, Application US/09428711A  
Patent No. 6358720  
GENERAL INFORMATION:  
APPLICANT: Muramatsu, Masaki  
APPLICANT: Shirasawa, Takuji  
APPLICANT: Tokumitsu, Hiroshi  
APPLICANT: NO. 6358720uchl, Teruhisa  
TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE  
FILE REFERENCE: 06501-045001  
CURRENT APPLICATION NUMBER: US/09/428,711A  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: PCT/JP98/01246  
PRIOR FILING DATE: 1998-03-23  
PRIOR APPLICATION NUMBER: JP 9/124798  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 3995  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (388) ... (3540)  
NAME/KEY: misc\_feature  
LOCATION: (1) ... (3995)  
OTHER INFORMATION: n = A,T,C or G  
US-09-428-711A-13

Query Match 2.4%; Score 37.2; DB 3; Length 3995;  
Best Local Similarity 46.2%; Pred. No. 12;  
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

```
QY 890 CCCCTACTGTGTGTGGTCTGAGGTACAGAGAACTGAGCTCTGTGGTGGCTTGGCG 949
DB 393 CTCATATAGCGCGCGGGGCGGAGCGGCTGGGCCCCGGGGCGGCTTACCG 334
QY 950 GGTGTGCGGAGATGATGAGAGTGTGCTGCTGTGTACTCATGTGCGCTGCTTCACTG 1009
DB 333 GGTGCGCTGGGTCTACGGCGGAGCGGCGGCGCGGCGCTGACGGAGACGAGCGCGGCTC 274
QY 1010 GGGCTGCCACTTCCAGCGGAGACTCCCGCGCGGAGCGGCGCTGCGGTGAGATCTCTG 1069
DB 273 ATGCGAGGGCGCGGTCCCATGTGGGAGCTGAGCGGGTCCAGTCCACAGCCCG 214
QY 1070 CTGAGAGAGCTGACCCGAGCCCTGTGAGAGGGGTGTGCTGCGCCCGAGCCCGGCT 1129
DB 213 GCGGACNAGAGGGGCGGCTCTCTCGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154
QY 1130 GGGCGCTGGGCT 1142
DB 153 GGACCGGTGACT 141
```

RESULT 90  
US-09-418-710-28  
Sequence 28, Application US/09418710  
Patent No. 6596482  
GENERAL INFORMATION:  
APPLICANT: Jones, Michael H.  
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR  
FILE REFERENCE: 06501-042001  
CURRENT APPLICATION NUMBER: US/09/418,710  
PRIOR FILING DATE: 1999-10-15  
PRIOR APPLICATION NUMBER: PCT/JP98/01783  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: JP 9/310027  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: JP 9/116570  
PRIOR FILING DATE: 1997-04-18  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 5561  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (346) ... (4926)  
US-09-418-710-28

Query Match 2.4%; Score 37.2; DB 4; Length 5561;  
Best Local Similarity 53.4%; Pred. No. 13;  
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
QY 535 ACGAGTGTGCGGTGTGTGCGGAGCGGAGAGCTATCTGTGTAGCGCTGCCCTGGG 594
DB 3908 AAGTTGTTCAGAAAGAGTGTGATCAATATTATCTTGTGTATAGTGTAAAG 3967
QY 595 CTTTCACCTGGCGCTGCTGCTCCCTCGCTCGGAGATCCCGAGTGGAGCTGAGGT 654
DB 3968 CTTTCACCTGTTTGTGTGTGAGCGCGGCTCTTATGAGTACAGATGTAGTGTGAGT 4027
QY 655 GCTTCAGCTGCTGAGGCAACAGTC 680
DB 4028 GCCGAGCTTGCAGCGCGCTACTGCC 4053
```

RESULT 91  
US-09-833-479-28  
Sequence 28, Application US/09839479  
Patent No. 6727222  
GENERAL INFORMATION:  
APPLICANT: Jones, Michael H.  
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR  
FILE REFERENCE: 06501-042002

```

; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 5561
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (346)...(4926)
US-09-839-479-28
```

```

Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5561;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```

QY 535 ACGAGTGTGCGCTGTGTCGAGACGGCGGAGCTCATCTGCTGTGACGGCTGCCCTCGGG 594
DB 3908 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTTGTTGATGATGAATTAAG 3967
QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACTGTGAGGT 654
DB 3968 CCTTCACCTGTTTGTCTGAGCGCCGCTCTATGATGACCAATGGAGTGGCAGT 4027
QY 655 GCTCCAGCTGCTGACGAGCAACAGTC 680
DB 4028 GCCCAGCTTGCCAGCCCGCTACTGCC 4053
```

```

RESULT 92
US-09-418-710-30
; Sequence 30, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 5573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (346)...(4938)
US-09-418-710-30
```

```

Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5573;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```

QY 535 ACGAGTGTGCGCTGTGTCGAGACGGCGGAGCTCATCTGCTGTGACGGCTGCCCTCGGG 594
DB 3920 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTTGTTGATGATGAATTAAG 3979
```

```

QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACTGTGAGGT 654
DB 3980 CCTTCACCTGTTTGTCTGAGCGCCGCTCTATGATGACCAATGGATGAGTGGCAGT 4039
QY 655 GCTCCAGCTGCTGACGAGCAACAGTC 680
DB 4040 GCCCAGCTTGCCAGCCCGCTACTGCC 4065
```

```

RESULT 93
US-09-839-479-30
; Sequence 30, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 5573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (346)...(4938)
US-09-839-479-30
```

```

Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5573;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```

QY 535 ACGAGTGTGCGCTGTGTCGAGACGGCGGAGCTCATCTGCTGTGACGGCTGCCCTCGGG 594
DB 3920 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTTGTTGATGATGAATTAAG 3979
QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACTGTGAGGT 654
DB 3980 CCTTCACCTGTTTGTCTGAGCGCCGCTCTATGATGACCAATGGATGAGTGGCAGT 4039
QY 655 GCTCCAGCTGCTGACGAGCAACAGTC 680
DB 4040 GCCCAGCTTGCCAGCCCGCTACTGCC 4065
```

```

RESULT 94
US-08-146-930-1/c
; Sequence 1, Application US/08146930
; Patent No. 5958764
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothenberg, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
```







QY 1095 GTGAGAGGGGTGTGCGCCCGCCAGCCCGCCGCTTGCGCCCTGAGGCTGCC 1145  
DB 114 CGCGCTTCGCGGCGACGCTGCTGCGCGGCTCTGTCGCGGTTCGCGGCGCACCC 64

## RESULT 98

US-09-022-765-34/C  
Sequence 34, Application US/09022765  
Patent No. 6375955  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,765  
FILING DATE: 12-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 516 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Leishmania major  
US-09-022-765-34

Query Match 2.4%; Score 37; DB 3; Length 516;  
Best Local Similarity 44.8%; Pred. No. 6.8;  
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGGCTTAAAGTCCGCGGAGAGAGGTAAGAGCTCCAGCCGGAACC 796  
DB 474 CGGCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGACGCTCTCGGCGTTCCGCGCACCT 415  
QY 797 CCTAGCGGCGATGACACGACTCTGTCTACAAGCAGCTGCGGCTCCGCTTCTGACAC 856  
DB 414 GCTGGGCGGCTCTGTCGCGGTCCGCGGCGACGCGTGTCCACCGCTGCTCAGCTCGGCC 355  
QY 857 CCCGCTGCGAGTCTGGAATCTCTGCGGCTTGCAACCCCTACTGTGTGTGAGTCTTGAAG 916  
DB 354 GCTGCTGCGGTGGCGGTGTCTAGGCGCTGCTGACGCTCTCGGCGTTCCGCGCACGCGTG 295  
QY 917 TCAGCAGAACT--GGCTCTGCTGCGGCTGCGGCTGCGGCTGCGGAGATGTAAGACGTC 974  
DB 294 CCACCTGTGCTCCACGCTGCGGCGGCTGCTGCGGCGGTGTCTAGGCGCTGCTGAGCT 235  
QY 975 CTGCGGTACTCACTGCGCGGCTGCTTCACTGCGGCTGCGCACTTCCAGCGGCGACC 1034

DB 234 CTTGCGGCTTCGCGGCGACGACTGTGACGCGGCTCTGCGGCTGCGGCGGCGGCGGCGCA 175  
QY 1035 TCCCGCGCGGAGCGGCGCTGCGCTGACAGATCTGCTCAGAGACGTACCCAGCCCT 1094  
DB 174 CCGGTGCTCCACGCTGCGGCGGCTGCTGCGGCGGTGTCTAGGCGCTGCTGACGCTCT 115  
QY 1095 GTGAGAGGGGTGTGCGCCCGCCAGCCCGCCGCTTGCGCCCTGAGGCTGCC 1145  
DB 114 CGCGCTTCGCGGCGACGCTGCTGCGCGGCTCTGTCGCGGTTCGCGGCGCACCC 64

## RESULT 99

US-09-551-974A-34/C  
Sequence 34, Application US/09551974A  
Patent No. 6500437  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C5  
CURRENT APPLICATION NUMBER: US/09/551,974A  
NUMBER OF SEQ ID NOS: 101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34  
LENGTH: 516  
TYPE: DNA  
ORGANISM: Leishmania major  
US-09-551-974A-34

Query Match 2.4%; Score 37; DB 4; Length 516;  
Best Local Similarity 44.8%; Pred. No. 6.8;  
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGGCTTAAAGTCCGCGGAGAGAGGTAAGAGCTCCAGCCGGAACC 796  
DB 474 CGGCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGACGCTCTCGGCGTTCCGCGCACCT 415  
QY 797 CCTAGCGGCGATGACACGACTCTGTCTACAAGCAGCTGCGGCTCCGCTTCTGACAC 856  
DB 414 GCTGGGCGGCTCTGTCGCGGTCCGCGGCGACGCGTGTCCACCGCTGCTCAGCTCGGCC 355  
QY 857 CCCGCTGCGAGTCTGGAATCTCTGCGGCTTGCAACCCCTACTGTGTGTGAGTCTTGAAG 916  
DB 354 GCTGCTGCGGTGGCGGTGTCTAGGCGCTGCTGACGCTCTCGGCGTTCCGCGCACGCGTG 295  
QY 917 TCAGCAGAACT--GGCTCTGCTGCGGCTGCGGCTGCGGCTGCGGAGATGTAAGACGTC 974  
DB 294 CCACCTGTGCTCCACGCTGCGGCGGCTGCTGCGGCGGTGTCTAGGCGCTGCTGAGCT 235  
QY 975 CTGCGGTACTCACTGCGCGGCTGCTTCACTGCGGCTGCGCACTTCCAGCGGCGACC 1034  
DB 234 CTTGCGGCTTCGCGGCGACGACTGTGCGGCGGCTCTGCGGCGGTGTCTAGGCGCTGCTGAGCT 175  
QY 1035 TCCCGCGCGGAGCGGCGCTGCGCTGACAGATCTGCTCAGAGACGTACCCAGCCCT 1094  
DB 174 CCGGTGCTCCACGCTGCGGCGGCTGCTGCGGCGGTGTCTAGGCGCTGCTGAGCTCT 115  
QY 1095 GTGAGAGGGGTGTGCGCCCGCCAGCCCGCCGCTTGCGCCCTGAGGCTGCC 1145  
DB 114 CGCGCTTCGCGGCGACGCTGCTGCGCGGCTCTGTCGCGGTTCGCGGCGCACCC 64

## RESULT 100

US-09-565-501A-34/C  
Sequence 34, Application US/09565501A  
Patent No. 6607731  
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skelky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Peter Probst  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
FILE REFERENCE: 210121.420C6  
CURRENT APPLICATION NUMBER: US/09/565,501A  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34  
LENGTH: 516  
TYPE: DNA  
ORGANISM: Leishmania major  
US-09-565-501A-34

Query Match 2.4%; Score 37; DB 4; Length 516;  
Best Local Similarity 44.8%; Pred. No. 6.8;  
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGGCTTAGTGGCGGAGAGAGTAAAGCTTCACTGGGGAGACC 796  
DB 474 CCGGCTGCTGCGTGGGCGTGTCTAGGCGCTGCTGCACTCTCGCGCTTCCGCGCAGCT 415  
QY 797 CCTAGCCGAGATGAGACGACTTGTCTAAGACACTGCGGCTCCGCGCTTCTGAGC 856  
DB 414 GCTGGCGGCTGCTGCGGCTCGCGGCAAGCCGCTGCAAGCTTCAAGCTCGGCGC 355  
QY 857 CCGGCTGCGAGGCTGAGCTCTCGGCTTGCACCCCTACTGTGTGTGAGG 916  
DB 354 GCTGCTGCGTGGGCGTGTCTAGGCGCTGCTGAGACTCTCGGCTTCCGCGCAGCG 295  
QY 917 TACGACAGAACTT--GGCTCTGTGCGCGCTTGGCGGAGTGTCCGAGATGTACGACGTG 974  
DB 294 CCACCTGTGCTTCAAGCTGCGGCGCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGCAAGCT 235  
QY 975 CTGCGGTGTACTCACTGCGCGCTTCACTGCGGCTGCGCACTTCCAGCCGAGCACC 1034  
DB 234 CTTGCGGCTTCCGCGGCAAGCTGCTGCGGCGCTTCTGCGCGTCCGCGCAGCGTCCA 175  
QY 1035 TCCCGGCTCGGAGACGCGGCTGCGCTGACAGATCTGCTCAAGAGACGTGACCCGAGCCCT 1094  
DB 174 CCGGTGCTTCAAGCTGCGGCGCTGCTGCGGCGGTGTCTAGGCGCTGCTGCAAGCTCT 115  
QY 1095 GTGAGAGGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1145  
DB 114 CGGCGTTCGCGGCAAGCTGCGGCGCTGCGGCGCTGCGGCGTCCGCGGCGCAGCC 64

Search completed: February 9, 2005, 14:36:14  
Job time : 195 sec

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 4527 Seconds  
(without alignments)  
12436.356 Million cell updates/sec

Title: US-09-508-658a-3

Predicted score: 1545  
Sequence: 1 agaagaagtgaggtcttc.....aaaataataaattgctg 1545

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 200 summaries

Database :

EST:\*  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_esc3:\*  
4: gb\_esc4:\*  
5: gb\_esc5:\*  
6: gb\_esc6:\*  
7: gb\_esc7:\*  
8: gb\_esc8:\*  
9: gb\_esc9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.4	39.6	642	5	BX112453 BX112453
2	542.6	35.1	1506	9	AY419550 Homo sapi
3	400	25.9	1474	9	AY419551 Pan trogl
4	374.6	24.2	433	1	AA742555 nx30c11.8
5	283.4	18.3	1512	9	AY419552 Mus muscu
6	162.6	10.5	451	2	BE627856 u49c07.y
7	156.2	10.1	632	6	BY728790 BY728790
8	149.4	9.7	1399	3	BC068026 Mus muscu
9	111.4	7.2	511	2	BE630816 u49c07.x
10	91.4	5.9	542	1	AA666822 vx91b07.r
11	90.2	5.8	420	1	CG513175 OST66421
12	86	5.6	384	1	AT552580 vx26b07.x
13	73.4	4.8	947	9	AG069825 Pan trogl
14	73	4.7	959	9	CNS035FO Droso
15	69.6	4.5	925	5	AL053013 Droso
16	69.4	4.5	674	5	BU740688 UI-E-EJO-
17	69.4	4.5	250	4	BM726504 UI-E-EJO-
18	67	4.3	940	7	CR566920 CR566920
19	65.6	4.2	350	8	AQ474914 CTTR1-EI-
20	65	4.2	563	1	AV591350 AV591350
21	64.2	4.2	610	2	BE590137 197111 BA
22	63.8	4.1	587	7	CNS322716 AGENCOURT
23	63.8	4.1	1203	9	CNS015Y4 AL106054 Droso
24	63.6	4.1	807	7	CK704772 ZP101-P00

25	63.6	4.1	888	9	AG030591 Pan trogl
26	63.4	4.1	783	5	BX915120 BX915120
27	63.2	4.1	406	1	AA792266 vn94e11.x
28	63.2	4.1	485	6	CF161888 B0703B06-
29	63.2	4.1	514	6	CA893864 B0181H02-
30	63.2	4.1	562	6	CA564994 R0332G02-
31	63.2	4.1	564	2	BB654405 BB654405
32	63.2	4.1	604	6	CD551263 B0325H01-
33	63.2	4.1	625	7	CK781630 UI-M-HBO-
34	63.2	4.1	635	5	BU705998 UI-M-F00-
35	63.2	4.1	635	5	BU055310 UI-M-F00-
36	63.2	4.1	723	6	CB246896 UI-M-F00-
37	63.2	4.1	730	7	CF726532 UI-M-HBO-
38	63.2	4.1	736	7	CNS34827 UI-M-HBO-
39	63.2	4.1	754	7	CO045219 UI-M-HO-
40	63.2	4.1	757	7	CNS36888 UI-M-HO-
41	63.2	4.1	762	6	CA327845 UI-M-FY0-
42	63.2	4.1	768	6	CA749488 UI-M-FY0-
43	63.2	4.1	787	7	CK637900 UI-M-HO-
44	63.2	4.1	799	6	CA752497 UI-M-F00-
45	63.2	4.1	832	5	BQ771341 UI-M-FIO-
46	63.2	4.1	855	5	BU152563 AGENCOURT
47	63.2	4.1	935	9	CNS0065XK Droso
48	63.2	4.1	1066	3	CR715927 Tetraodon
49	63.2	4.1	3337	3	AK034549 Mus muscu
50	62.8	4.1	925	9	CNS0091P Droso
51	62.6	4.1	587	4	BG511885 dad29d08.
52	62.2	4.0	677	7	CNS26117 UI-M-HNO-
53	62.2	4.0	680	6	CB523529 UI-M-GHO-
54	62.2	4.0	751	5	BU056500 UI-M-F00-
55	61.8	4.0	776	9	AG561783 Mus muscu
56	61.6	4.0	262	9	CE173623 t1gr-g8a
57	61.4	4.0	434	2	BE234617 141780 MA
58	61.4	4.0	670	7	CF792818 885797 MA
59	61.4	4.0	672	7	CF794185 889669 MA
60	61.4	4.0	826	7	CNS153741 940889 MA
61	61.4	3.9	520	4	BG307509 f159810.x
62	60.8	3.9	429	5	BY389139 BT389139
63	60.8	3.9	804	7	CNS35140 UI-M-HSO-
64	60.4	3.9	935	9	CNS0065XK Droso
65	60	3.9	359	2	BE814801 QV4-BN009
66	60	3.9	461	6	CF137810 UI-HF-BNO
67	60	3.9	544	5	BU429662 UI-HF-BNO
68	60	3.9	593	7	CN483764 h37c03.y
69	60	3.9	683	1	AU127678 AU127678
70	60	3.9	700	4	BE238675 602429276
71	60	3.9	731	2	BE176589 RC3-HT058
72	60	3.9	750	2	BE297517 601178063
73	60	3.9	817	2	BE259148 601108075
74	60	3.9	878	5	BQ437443 AGENCOURT
75	60	3.9	948	2	BE299945 600944490
76	60	3.9	974	4	BG394104 602456272
77	60	3.9	989	4	BM457082 AGENCOURT
78	60	3.9	989	4	BM457082 AGENCOURT
79	60	3.9	1011	2	BF309509 601892016
80	60	3.9	1365	5	BQ213031 AGENCOURT
81	59.8	3.9	426	2	BF487408 PM4-EN006
82	59.8	3.9	885	5	BU193516 AGENCOURT
83	59.8	3.9	918	5	BU527069 AGENCOURT
84	59.8	3.9	943	5	BU228543 AGENCOURT
85	59.4	3.8	1104	5	BU722521 AGENCOURT
86	58.8	3.8	1132	5	BQ672163 AGENCOURT
87	58.6	3.8	1350	4	BM450237 AGENCOURT
88	58.4	3.8	416	6	CB606244 AMGNNUC:S
89	58.4	3.8	932	5	CNS00720 Droso
90	58.2	3.8	302	9	BP761124 BP761124
91	58.2	3.8	357	7	CF535776 UI-M-GHO-
92	58.2	3.8	361	7	CF535776 UI-M-GHO-
93	58.2	3.8	428	1	AA839768 v51d11.x
94	58.2	3.8	452	6	CB520637 UI-M-GIO-
95	58.2	3.8	575	7	CF911441 A0615D04-
96	58.2	3.8	690	5	BU703948 UI-M-F00-
97	58.2	3.8	2253	3	AK045449 Mus muscu

98	58	3.8	932	9	CNS00720	AL066742	Drosophila
99	58	3.8	997	2	AG057392	AG057392	Pan trogl
100	57.8	3.7	1038	5	BQ953925	BQ953925	AGENCOURT
C 101	57.6	3.7	1063	5	CL498639	SAILL_657	
102	57.4	3.7	745	7	CK693917	2F101-P00	
103	57.2	3.7	251	5	BU056338	UI-M-F00-	
104	57.2	3.7	446	7	CF537874	UI-M-F00-	
105	57.2	3.7	1094	7	CK161715	FGAS01428	
106	57	3.7	993	2	BE874434	601488870	
C 107	56.8	3.7	1148	9	AG346700	Mus muscu	
C 108	56.8	3.7	1359	9	CL507830	SAILL_788	
C 109	56.6	3.7	459	6	CF143165	UI-HF-BRO	
110	56.6	3.7	540	2	AM502232	UI-HF-BRO	
111	56.6	3.7	862	4	BG758185	602712357	
112	56.6	3.7	934	5	BX410043	BX410043	
C 113	56.4	3.7	1523	9	AG448267	Mus muscu	
114	56.2	3.6	208	6	CF182368	UI-M-EY0-	
115	56.2	3.6	400	1	AJ683050	AJ683050	
116	56.2	3.6	579	7	CN482597	hw22c09.Y	
117	56.2	3.6	740	7	CF539286	UI-M-G10-	
118	56.2	3.6	798	6	CD352609	UI-M-G10-	
119	56.2	3.6	820	1	AU118164	AU118164	
C 120	56	3.6	665	9	AG159109	Pan trogl	
C 121	56	3.6	1355	9	AG429718	Mus muscu	
122	55.8	3.6	452	2	BF392580	UI-R-CAO-	
123	55.8	3.6	535	2	BF416810	UI-R-CAO-	
124	55.8	3.6	537	2	BF412941	UI-R-CAO-	
125	55.8	3.6	547	2	BE115719	UI-R-CAO-	
126	55.8	3.6	550	2	BF394139	UI-R-CAO-	
127	55.8	3.6	554	2	BF418125	UI-R-CAO-	
128	55.8	3.6	567	2	BE107720	UI-R-CAO-	
129	55.8	3.6	570	2	BF416261	UI-R-CAO-	
C 130	55.8	3.6	1462	9	AG441877	Mus muscu	
131	55.6	3.6	540	2	BF899692	QV1-MT022	
132	55.6	3.6	957	9	AG077010	Pan trogl	
133	55.6	3.6	1070	6	CA987146	AGENCOURT	
134	55.6	3.6	1312	4	BM914111	AGENCOURT	
135	55.4	3.6	1853	9	AG382854	Mus muscu	
C 136	55.4	3.6	1890	9	AG346134	Mus muscu	
137	55.2	3.6	392	6	CA876606	K0949F05-	
138	55.2	3.6	426	6	CA871588	K0911B08-	
139	55.2	3.6	537	1	AI036739	uc90a11.Y	
140	55.2	3.6	566	7	CN688217	BE0254G07-	
141	55.2	3.6	646	2	BS858201	BS858201	
C 142	55.2	3.6	1203	9	CNS01574	AG441877	Mus muscu
143	55.2	3.6	1462	9	AG441877	Mus muscu	
144	55	3.6	412	2	BE018328	BE018328	
145	55	3.6	872	9	AG435681	AG435681	
C 146	55	3.6	1961	9	AG435681	AG435681	
147	54.8	3.5	485	9	CE447963	CE447963	
148	54.8	3.5	485	9	CE447963	CE447963	
149	54.8	3.5	899	9	CC700175	OGLB018TH	
C 150	54.8	3.5	965	8	AQ893417	HS_4832_A	
151	54.8	3.5	1089	5	BQ648572	AGENCOURT	
152	54.8	3.5	1280	9	CL470317	SAILL_140	
153	54.8	3.5	1552	9	AG430101	Mus muscu	
154	54.6	3.5	420	1	AL120275	DKFZP61L	
155	54.6	3.5	457	2	BB852481	BB852481	
156	54.6	3.5	502	2	BF286221	EST450812	
157	54.6	3.5	531	2	BF899690	QV1-MT022	
C 158	54.6	3.5	583	4	BG375097	UI-R-CV1-	
159	54.6	3.5	645	4	B1380872	BFLG1_002	
160	54.6	3.5	720	7	CO400030	AGENCOURT	
161	54.6	3.5	842	9	AG058791	Pan trogl	
C 162	54.6	3.5	1101	9	CNS0178Y	AL108460	Drosophila
C 163	54.6	3.5	1628	9	CG757066	P052-2-A0	
C 164	54.4	3.5	519	2	BF899766	QV1-MT022	
C 165	54.4	3.5	1112	6	CA477644	AGENCOURT	
166	54.2	3.5	284	5	BQ436154	AGENCOURT	
C 167	54.2	3.5	301	2	BF800390	IL5-CI000	
C 168	54.2	3.5	308	2	AW513912	xc90d06.x	
C 169	54.2	3.5	342	7	T05980	EST03869	Fe
170	54.2	3.5	529	7	CN410758	170005315	

171	54.2	3.5	618	4	BG574441	BG574441	602596388
172	54.2	3.5	689	2	AM961105	AM961105	EST373072
173	54.2	3.5	689	1	AV707144	AV707144	
174	54.2	3.5	701	4	BM788417	K-EST0067	
175	54.2	3.5	822	5	BU568076	BU568076	AGENCOURT
176	54.2	3.5	824	4	BG565128	602583752	
177	54.2	3.5	870	4	BG614027	602641565	
C 178	54	3.5	1569	9	AG341503	Mus muscu	
C 179	53.8	3.5	742	9	CG696734	MUGQ_CH25	
180	53.8	3.5	809	4	BI259307	602972544	
181	53.8	3.5	1265	9	CG750734	P045-2-B0	
182	53.8	3.5	1314	9	AG039591	Pan trogl	
183	53.6	3.5	249	5	BQ308190	MRO-BT450	
C 184	53.6	3.5	611	4	BI044950	MR4-OT010	
185	53.6	3.5	662	7	CN277795	170005313	
C 186	53.6	3.5	726	9	BX174921	Danio rer	
187	53.6	3.5	853	5	BUB55947	AGENCOURT	
188	53.6	3.5	924	4	BG395346	602457885	
189	53.6	3.5	1004	5	BU149386	AGENCOURT	
C 190	53.6	3.5	1137	4	BG809979	mgc002xd	
C 191	53.6	3.5	1830	9	CL078620	CH216-151	
C 192	53.6	3.5	2243	9	AG381986	Mus muscu	
C 193	53.4	3.5	428	5	BX916095	BX916095	
C 194	53.4	3.5	428	5	BX922861	BX922861	
195	53.4	3.5	949	9	AG171092	Pan trogl	
196	53.4	3.5	1452	9	AG032979	Pan trogl	
197	53	3.4	738	5	BU326843	603492618	
198	53	3.4	1088	4	BM450730	AGENCOURT	
199	53	3.4	1093	4	BM803558	AGENCOURT	
C 200	53	3.4	1160	9	AG043473	Pan trogl	

## ALIGNMENTS

RESULT 1  
LOCUS BX112453  
DEFINITION BX112453 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1257620, mRNA sequence.  
KEYWORDS EST.  
ACCESSION BX112453  
VERSION BX112453.1 GI:27837617  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M., Radloff, U., Schneider, D. and Korn, B.  
TITLE Human Unigeneset - RZPD3  
JOURNAL Unpublished (2003)  
COMMENT Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD: IMAGP98F231315.  
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;  
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi?responsefileNo=972  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available royalty-free from RZPD;  
Contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTCACACAGCAACAGCTATGAC.  
Location/Qualifiers

## FEATURES

source  
1..642  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"



[illegible][illegible]





OY		1112	CCCCAGCCCGCCGCGCTTGAGGCCCTTCGGAGCCTGCCAAGATACACCTGCATCTCAACGACC	1171
Db		1332	NNNACA	1391
OY		1172	CGCTTCGACACGGGATGACTGAGTCCCTTTCTGAGCGAGACA	1231
Db		1392	TGTTTACATATGAGGAAGAAGCCTGGAGATCCCTCCTCAATGAGCATCTATTGACGGCATCCT	1451
OY		1232	GCACTGGGCGCATCCAGACATGAGCCCTTCGGAGCCGCCCTTCCCTCT	1281
Db		1452	GCAGTGGGCGCATTCAGACATGTACGCGCCGCTGGCGAGACACACCT	1501
RESULT 6				
LOCUS	BE627856			
DEFINITION	BE627856	451 bp	mRNA	linear EST 24-AUG-2000
	sequence.	uuagcctg.y1 Soares_thymus_2dbmt Mus musculus cDNA clone		
IMAGE:	3375276 5'	similar to TR:Q920E3 Q920E3 AIRE PROTEIN.		
ACCESSION	BE627856			
VERSION	BE627856.1	GI:9908250		
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
TITLE	1 (bases 1 to 451) NCI-GARP http://www.ncbi.nlm.nih.gov/hicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)			
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs.r@mail.nih.gov This clone is available royalty-free through LNL ; contact the IMGE Consortium (info@image.lnl.gov) for further information. MGI:1084880			
COMMENT	Seq primer: -40RP from Gibco High quality sequence stop: 356. Location/Qualifiers			
FEATURES	source	1..451		
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:3375276"			
	/sex="male"			
	/tissue_type="Thymus"			
	/dev_stage="4 weeks"			
	/lab_host="DH10B"			
	/clone_lib="Soares_thymus_2dbmt"			
	/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTTCACATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."			
ORIGIN				
Query Match	10.5%; Score 162.6; DB 2; Length 451;			
Best Local Similarity	71.8%; Pred. No. 5.5e-26;			
Matches	245; Conservative 0; Mismatches 84; Indels 12; Gaps 2;			
OY	953	GTGCGAATGTGTAGGACGTGCTCGGATGTACTACATGCGCGCTGCTTCACTGAGC	1012	
Db	1	GTGTGGCATGTGACCGAGAGTGTTCGGGTGTGACAACTGTGCGCTGCTTCACTGAGC	60	
OY	1013	CTGCCACTTCCAGCGCGGACCTCCCGGCGCGGAGCGGAGCTGCGCTGCAAGTCTGCTC	1072	
Db	61	CTGCCACTTCCAGCGCGCGCGCGCGCGCGGAGCAATCTCCGCTGCAATCTGCTC	120	

RESULT 7	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang L., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N., Hizoane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Satoh K., Shiraki T., Waki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang L., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N., Hizoane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Satoh K., Shiraki T., Waki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang L., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N., Hizoane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Satoh K., Shiraki T., Waki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang L., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N., Hizoane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Satoh K., Shiraki T., Waki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang L., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N., Hizoane-Kishikawa T., Komno H., Nakamura M., Sakazume N., Satoh K., Shiraki T., Waki K., Kawai J., Aizawa K., Aizawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E. and Hayashizaki Y.
Oy	BY728790	632 bp mRNA linear EST 17-DEC-2002	BY728790	1	GI:27141917	Mus musculus (house mouse)	Mus musculus	1	Okazaki Y., Furuno M., Kaubawa T., Adachi J., Bono H., Kondo S., Nikaide I., Oseto N., Saito R., Suzuki H., Yamana K., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schondach C., Gotohori T., Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Frazier K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guatlichin S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,

Imorani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ono, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct Submissions

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

## source

Location/Qualifiers  
1..632  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="C330016003"  
/cell\_type="ES cells"  
/lab\_host="SOLR"  
/clone\_lib="RIKEN full-length enriched, ES cells"  
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']  
GAGAGAGAGATCTCGAGTTAATTAATTCACCCGCCCC 3'."

## ORIGIN

Query Match 10.1%; Score 156.2; DB 6; Length 632;  
Best Local Similarity 69.4%; Prid. No. 1.6e-24;  
Matches 240; Conservative 0; Mismatches 104; Indels 2; Gaps 2;  
Db 502 TGGGGCTCTTGGCCGTGCGAGAGATGAGGAGTGGCGGTGCGGAGGGG 561  
251 TGGGTCCTCTTGCTG-CTCAGAGAGACAGATGATGTCGCTGTCACAGAG 309  
562 GGGAGCTACTGCTGTCGAGCGCTGCTGCGCTTCCACCTGCTGCTCTCC 621  
310 GTGAGCTACTGCTGTCGAGCGCTGCTGCGCTTCCACCTGCTGCTCTCC 369  
622 CGTCCGGAGATCCCGAGTGGAGCTGTCAGCTGTCGAGGCAACAGTCC 681  
370 CTCTGAGAGAGATCCCGAGTGGCTGTCGAGTCTCTGCTCCAGGAGAGTCC 429  
682 AGAGGTGAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 741  
430 AACAGAACTGTCTCCAGCTGAGAGTCTCCAGGCGCCCGAGAGTAACTGCA 489  
742 TCCCGCCGGGGCTTAGTGGCGGAGAGAGAGTAAAGAGTCCACTGGGGAAC 801  
490 ACCAATCTCGGCTT-GCAATCTGCTCTGCAAGCTGAGTCCAGCCAGGAGCAG 548

QY 802 CCGGCTGAGACGACTCTTGTCTACAGCACCCTCGCGCTCCGCC 847  
Db 549 CCAAGCTGACACCTCTGGGCGCCGCTCCAGCACCCTGGGCTTGCC 594

## RESULT 8

LOCUS BC068026 1399 bp mRNA linear HTC 25-MAR-2004  
DEFINITION Mus musculus cDNA clone IMAGE:1282549, containing frame-shift errors.

ACCESSION BC068026  
VERSION BC068026.1 GI:45751562

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

1 (bases 1 to 1399)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.P., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marsina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullan, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hu, X., Gibbe, R.A., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbe, R.A., Paney, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Scheraga, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1399)  
Strausberg, R.  
Direct Submission  
Submitted (24-MAR-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ni.nih.gov>  
Contract: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Marcello Bento Soares, Ph.D.

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

BC Cancer Agency, Vancouver, BC, Canada

Info@gscc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,

Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth

Featherstone, Malachi Griffith, Ohi Griffith, Ran Guin, Nancy Liao,

Kim Macdonald, Amara Maeson, Mike R. Mayo, Josh Moran, Ryan Morin,

Telka Olson, Diana Palmquist, Anca Petrescu, Anna Lilia Prabu,

Parvaneh Saeedi, JR Santos, Angeliqne Scherch, Ursula Skalska,

Duane Smalley, Jeff Stolt, Miranda Tsai, George Yang, Jacques

Schein, Asim Siddiqui, Rob Holt, Marco Marra.

## FEATURES

## source

1..1399  
/organism="Mus musculus"

	/mol_type="mRNA"
	/strain="CS7BL/6J"
	/db_xref="taxon:10090"
	/clone="IMAGE:1282549"
	/issue_type="Thymus gland, mouse"
	/clone_id="Soares_thymus_2NBWT"
	/lab_host="DH10B"
	/note="Vector: pRTT3-Pac"

  

Query Match	9.7%	Score 149.4;	DB 3;	Length 1399;
Best Local Similarity	79.4%;	Pred. No. 6e-23;		
Matches 177;	Conservative 0;	Mismatches 46;	Indels 0;	Gaps 0;
OY	526 AGAATGAGGACGAGTGTCCTCGGTGTGTGGGACGCCGGGGAAGTCAATTCTGTGTAGCCGT	585		
Dd	442 AGAACGAGAATGAGTGTCCTCGGTGTGCACGACGAGGTGAGACTCATCTGTGTAGCAGCT	501		
OY	586 GCCCTCGGGCCCTTCACCTGGCGTGCCTCGTCCCCTCGCTCGGGAGATCCCCAGTGGGA	645		
Dd	502 GTCCCCGGGACCTTCACCTGGGTGCTGTGCTCCACCCTTCGACGAGAGATCCCAGTGACC	561		
OY	646 CTTGAGGAGTGCTCCAGCTGCTGTGACGACAAGTCCAGAGATGTCAGCCCGGCGACAGG	705		
Dd	562 TCTGGAGATGCTCTCGTCTGCTCCAGGGSCAGAGTCCAACAATACTGTGCCAGCTAGAG	621		
OY	706 AGCCCCGGGCCCGAGAGCACCCCGTGTGAGACCCCGCTCCCCC	748		
Dd	622 TGTCAGAGCCCCCGAGCTACCTGTGACAGAGACCCCGGTATGCC	664		

  

RESULT 9	BE630816/c	511 bp	mRNA	linear	EST 25-AUG-2000
LOCUS	BE630816				
DEFINITION	uu49c07.x1 Soares thymus_2NBWT Mus musculus CDNA clone				
VERSION	IMAGE:3375276 3'				
KEYWORDS	similar to TR:Q9Z0E3 Q9Z0E3 AIRE PROTEIN.				; mRNA sequence.
ACCESSION	BE630816				
VERSION	BE630816.1	GI:9913504			
SOURCE	EST.				
ORGANISM	Mus musculus (house mouse)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
AUTHORS	I (baaes i to 511)				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL	Tumor Gene Index				
COMMENT	Unpublished (1997)				
	Other_BSTRs: uu49c07.y1				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	This clone is available royalty-free through LNL / contact the				
	IMAGE Consortium (info@image.llnl.gov) for further information.				
	MGJ:1084880				
FEATURES	High quality sequence stop: 331.				
SOURCE	Location/Qualifiers				
	1..511				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="CS7BL/6J"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:3375276"				
	/sex="male"				
	/tissue_type="Thymus"				
	/dev_stage="4 weeks"				
	/lab_host="DH10B"				
	/clone_id="Soares_thymus_2NBWT"				
	(note=)Vector: pRTT3-Pac (Pharmacia) with a modified				
	polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA				
	was primed with a Not I - oligo(dTp) primer [5'				
	TGTTACCAATCTGAAGTAGGAGCGCCGCTTTTTTTTTTTTTTTTTTTT				
	3']; double-stranded cDNA was ligated to Eco RI adapters				

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7n3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AA866822	542 bp mRNA linear EST 16-MAR-1998	AA866822									
	<p>U3107.71 Soares_thymus.2NDMT Mus musculus cDNA clone IMAGE:1282549 5' similar to TR:015164 015164 TRANSCRIPTION INTERMEDIARY FACTOR 1. ;, mRNA sequence.</p> <p>AA866822</p> <p>AA866822.1 GI:2962267</p> <p>EST.</p> <p>Mus musculus (house mouse)</p> <p>Mus musculus</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.</p> <p>1 (bases 1 to 542)</p> <p>Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisell,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.</p> <p>The Mashu-HHMI Mouse EST Project</p> <p>Unpublished (1996)</p> <p>Contact: Marra M/Mouse EST Project</p> <p>Washu-HHMI Mouse EST Project</p> <p>Washington University School of Medicine</p> <p>444 Forest Park Parkway, Box 8501, St. Louis, MO 63108</p> <p>Tel: 314 286 1800</p> <p>Fax: 314 286 1810</p> <p>Email: mouseest@watson.wustl.edu</p> <p>This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.</p> <p>MGI:674349</p> <p>Possible reversed clone: similarity on wrong strand</p> <p>Seq primer: -28ml3 rev2 RT from Amersham</p> <p>High quality sequence stop: 472.</p> <p>Location/Qualifiers</p> <p>1..542</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:1282549"</p>										





MEDLINE	20296633
PUBMED	10835645
REFERENCE	2
AUTHORS	Roeck Crolius,H., Jallou,O., Basilya,C., Ozouf-Costaz,C., Plamies,C., Fischer,C., Bouneau,D., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000)
TITLE	
JOURNAL	
MEDLINE	20359837
PUBMED	10899143
REFERENCE	3 (bases 1 to 959)
AUTHORS	
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) This sequence is a single read and was generated as part of a large scale clone-and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .
COMMENT	
FEATURES	location/Qualifiers
SOURCE	1..959 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone="022016" /clone_1lb="G" /note="Genoscope sequence ID : COBG022BH08SP1-end : PUC-Ori"
ORIGIN	
Query Match	4.7% ; Score 73; DB 9; Length 959;
Best Local Similarity	60.8% ; Pred.No.1.le-05;
Matches 110; Conservative	5; Mismatches 66; Indels 0; Gaps 0;
Oy	522 CAGAAGATTAGAGACAGACTGTCTGTCTGGACCGCGGAGAGCTCATCTGCTGTAC 581
Db	434 CTGGAGCAATGAGAGACAATGTGCCCTTCGCAGAGAGACGAGACTCAGCCCTGCGCG 493
Oy	582 GGCTTGCCCTGGGGGCTTCACCTGCGCTGCGCTGCTCCCTCCGCTCCGGAGATCCCACT 641
Db	494 AGCTGCTVCCCCGCTTMAACCCCAAGCTGCTGTGACCCCCCAMTAATAAACCCCCCGG 553
Oy	642 GGAAGCTGAGAGGTCTCCAGCTGCTGCTGCAAGGCAACAGTCCAGAGAGTCCAGCCCCGGACA 701
Db	554 GGCCCTGTGTTCTGCCCAAGTGCAGAAAGTAATTCTGTGCTCCCACTAATCTCA 613
Oy	702 G 702
Db	614 G 614
RESULT 15	
LOCUS	CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACK19D16 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1 GI:4934461
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyndroidea; Drosophilidae; Drosophila.
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library](http://www.fruitfly.org/The%20BDGP/Drosophila%20melanogaster/BAC%20library.html) was prepared by Kazuo Oosawa and Aaron Mammone in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES             Source
    location/Qualifiers
        . 925
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BACR19D1c"
            /clone_1kb="RPc1-98"
            /note="End : TRJ3"

ORIGIN
Query Match          4.5%; Score 69.6; DB 9; Length 925;
Best Local Similarity 13.5%; Pred. No. 6.2e-05;
Matches 46; Conservative 173; Mismatches 123; Indels 0; Gaps 0;

Oy      833 CTTGCGGGCTCCGCTTGTGCAGCCCCGCTGCCAGATCTTGACCTTCGGCCTTGACACC 892
         |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       575 CSSCSGSSSSSCCBCCCCSSSYCCSSBSBKSTBSGSCCSSSKSVCTGSSS 634
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      893 CCTACTGTGTGCGTCTGTAGAGTCAGAACTGATCGCTTCGTGCGCGTGGGGGT 952
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       635 SSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 694
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      953 GTGCGGAGATGTACGACAGCTGCGCGTACTACTGCGCCGCTGCTTCACACTGACC 1012
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       695 TSSSSSSSSSTSSSSSVSGKSTBSGSSSSSSSSSSSSSSSSSBCTSTSSSSSSSS 754
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      1013 CTGCACTTCCCACCGGACCTCCGAGCGGAGCGGCGCTGCGGTCAGATCTGCTC 1072
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       755 SSTSCTCCCSYSSSSTSSSSTSMGSTBSGSSSSSVGTSSSSDSTSTCCSCCTWCTC 814
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      1073 AGAGAAGTAGACCCAGCCCCCTGTGAAGAGGGGTGCTGCCCCAGCCCGCGCTGAC 1132
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       815 CSTYMBGCTSTSCGSSSSGKGVTAKCGCGCGSSTNGMGTSTACSSSSSSSSSS 874
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      1133 CCCTGGGCGCTGCCAAGATGACACTGCCCATGACAGACCCGC 1174
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       875 VSSSSKSSASSSVSSSGSGVSSSSSSASKSsgsvssgs 916
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 16
BU740688/c 674 bp mRNA linear EST 10-OCT-2002
DEFINITION UI-E-BJ0-a11-m-23-0-UI.a1 UI-E-BJ0 Homo sapiens cDNA clone
ACCESSION U1-E-BJ0-a11-m-23-0-UI 3', mRNA sequence.
VERSION BU740688
KEYWORDS BU740688.1 GI:23680847
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
EXTRACT Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 674)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINS 97044477
PUBMED 8889548
```



## COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
The following repetitive elements were found in this CDNA  
sequence: 405-484 >(GGA)n#simple\_repeat (matched complement)  
Seq primer: M13 FORWARD  
POLYA=yes.

## FEATURES

source

Location/Qualifiers  
1..674  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-EJ0-41-m-23-0-UI"  
/tissue\_type="fetal eyes, lens, eye anterior segment,  
optic nerve, retina, Retina Foveal and Macular, RPE and  
Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-EJ0"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-EJ0 is a subcloned CDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pT7T3-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are: fetal eyes,  
AAATCAAGA; lens, CGATTAGCA; eye anterior segment,  
AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG;  
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This  
library was created for the program, Gene Discovery in the  
Visual System, supported by National Eye Institute (NEI).  
TAG\_TISSUE=human fetal eyes  
TAG\_LIB=UI-E-EJ0  
TAG\_SEQ=AAATCAAGA"

## ORIGIN

Query Match 4.5%; Score 69.4; DB 5; Length 674;  
Best Local Similarity 60.3%; Pred. No. 6.6e-05;  
Matches 112; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY CTCTTGCTGTGCGCAGAGATGAGGACGAGTGTGTCGGGACGCGCGGAGCT 568  
DB CTGCTGTCTGTCTCAAGCGCTGATTAATCTGTGAGTGTGCGACAGGATGAGAT 582  
QY CATCTGCTGTGACGCGTGTGCGGCTCTTCCACTGCGCTGTCTCTCCGCTCCG 628  
DB CATCTGCTGTGACGACCTGCGCGGCGCTACCATCTCTGTGCTGTGAGACCGAGCTGGA 522  
QY GGAGATCCCGAGTGGGACCTGAGGCTGCCAGCTCCCGTCCGCGGACCAAGTCCAGAGGT 688  
DB GATGGCTCCCGAGGAGGAGTGGAGCTGCCCGCACTGTGAGAAAGAGGAGATCCAGTGGGA 462  
QY GCAG 692  
DB GCGG 458

RESULT 17

## BM726504

LOCUS BM726504 250 bp mRNA linear EST 01-MAR-2002  
DEFINITION UI-E-EJ0-41-m-23-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone  
ACCESSION UI-E-EJ0-41-m-23-0-UI.5, mRNA sequence.  
VERSION BM726504  
KEYWORDS BM726504.1 GI:19047837  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 250)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery

MEDLINE Genome Res. 6 (9), 791-806 (1996)

PUBMED 97044477

## COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

## FEATURES

source

Location/Qualifiers  
1..250  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-EJ0-41-m-23-0-UI"  
/tissue\_type="fetal eyes, lens, eye anterior segment,  
optic nerve, retina, Retina Foveal and Macular, RPE and  
Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-EJ0"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-EJ0 is a subcloned CDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pT7T3-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are: fetal eyes,  
AGATCAAGA; lens, CGATTAGCA; eye anterior segment,  
AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG;  
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This  
library was created for the program, Gene Discovery in the  
Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Query Match 4.5%; Score 69; DB 4; Length 250;  
Best Local Similarity 61.3%; Pred. No. 7.3e-05;  
Matches 111; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 512 TTGCTGTGCGCAGAGATGAGGACGAGTGTGTCGGGACGCGCGGAGCTCAT 571  
DB 37 TGTCTGTCTCAAGCGCTGATTAATCTGTGAGTGTGCGACAGGATGAGATCAT 96  
QY 572 CTGCTGTGACGCGCTGCGGCTCTTCCACTGCGCTGTCTCTCCGCTCCGGA 631

Db 97 CCTGTGCGACACTCTGCCCGGACCTACCTCTGTGCTGCTGACCCAGAGCTGGAGAA 156  
 QY 632 GATCCCAAGTGGAGCTGGAGTGTCTCAAGCTGCTTCCAGCAACAGTCCAGAGGTGCA 691  
 Db 157 GGCTCCCGAGGAGCAAGTGGAGTGTGCCCCCACTGTGAGAGAGGAGTCCAGTGGAGGCC 216  
 QY 692 G 692  
 Db 217 G 217

RESULT 18  
 CR566920 940 bp mRNA linear EST 19-JUL-2004  
 LOCUS CR566920 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA010e06  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CR566920  
 VERSION CR566920.1 GI:50396997  
 KEYWORDS EST.  
 SOURCE Xenopus tropicalis (western clawed frog)  
 ORGANISM Xenopus tropicalis  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus; Silurana.  
 1 (bases 1 to 940)  
 Crouching, M.D.R., Ashhurst, J.L., Taylor, R., Garrett, N. and Rogers, J.  
 Sanger Xenopus tropicalis EST project 2001 (2004)  
 Unpublished (2004)  
 COMMENT Contact: Crouching MDR  
 Sanger Institute  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS SEQUENCE ID: THDA010e06.pikbSp6  
 This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Nigel Garrett.  
 Seq primer: Sp6.  
 Location/Qualifiers  
 1. 940  
 /organism="Xenopus tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="THDA010e06"  
 /dev\_stage="tailbud head (stage 28-30)"  
 /lab\_host="Escherichia coli DH10B."  
 /clone\_lib="XGC-tailbud-head"  
 /note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
 was oligo dt primed from 5ug of poly A+ RNA from tailbud  
 head. EcoRI-NotI cut cDNA was then ligated into PCS107  
 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN  
 Query Match 4.3%; Score 67; DB 7; Length 940;  
 Best Local Similarity 66.0%; Pred. No. 0.00024;  
 Matches 97; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 519 TGGCAGAGAGATGAGACGAGTGTGCTGTGCGAGCGCGGAGACTCATCTGCTGT 578  
 Db 245 TACCAAGATATGATGATGATGCTCTGCTGCTGAGAGATGGGGGAGTAAATATGTTGC 304  
 QY 579 GACGGTGTCTGCTGGGCTTTCCACCTGAGCTGTGCTTCCCTCCGCTCCGGAGATCCCC 638  
 Db 305 GATGAGATGCCCAAGGCTTTTCACTTCTGCTGTGCTGGCCGCTTTGACCCATATTTCCA 364  
 QY 639 AGTGGACCTGAGGTGCTTCCAGCTGC 665  
 Db 365 AGCGGACATGAGATGTGATCTTGC 391

RESULT 19  
 A0474914 350 bp DNA linear GSS 23-APR-1999  
 LOCUS A0474914  
 DEFINITION CITBI-EI-2591L22.TF CITBI-EI Homo sapiens genomic clone 2591L22,

genomic survey sequence.  
 A0474914  
 VERSION A0474914.1 GI:4657033  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 350)  
 Zhao, S., Adams, M.D., Niernman, W., Malek, J., Shizuya, H., Simon, M. and  
 Venter, J.C.  
 Venter, J.C.  
 TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other GSSs: CITBI-EI-2591L22.TF  
 Contact: Shaying Zhao, William Niernman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are available from Research Genetics (info@reggen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13-21  
 Class: BAC ends.  
 Location/Qualifiers  
 1. 350  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="2591L22"  
 /sex="male"  
 /cell\_type="sperm"  
 /clone\_lib="CITBI-EI"  
 /note="Vector: pBelOBAC11; Site 1: EcoRI; Site 2: EcoRI;  
 Caltech Human BAC Library D"

ORIGIN  
 Query Match 4.2%; Score 65.6; DB 8; Length 350;  
 Best Local Similarity 94.4%; Pred. No. 0.00044;  
 Matches 68; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1145 CAGAGTACACTGCTCCAGTACAGCCGCTCTGACAGGATGACTGAGTCCCTTCT 1204  
 Db 109 CCAGATGACACTGCGACAGTCAAGCCGCTCTGACAGGATGACTGAGTCCCTTCT 168  
 QY 1205 GAGCGAGCACAC 1216  
 Db 169 GAGCGAGGTAC 180

RESULT 20  
 A0474914 563 bp mRNA linear EST 27-NOV-2001  
 LOCUS A0474914 Bos taurus brain fetus Bos taurus cDNA clone E18018A07  
 DEFINITION 5', mRNA sequence.  
 ACCESSION A0474914  
 VERSION A0474914.1 GI:9702343  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 563)  
 Takasuga, A., Hirotsune, S., Itoh, R., Itoh, R., Suzuki, H., Aso, H.  
 and Sugimoto, Y.  
 Establishment of a high throughput EST sequencing system using  
 poly(A) tail-removed cDNA libraries and determination of 36,000  
 bovine ESTs  
 Nucleic Acids Res. 29 (22), E108 (2001)

**MEDLINE** 21570554  
**PUBMED** 11713328  
**COMMENT** Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-Shitakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazuugi@cocoa.ocn.ne.jp  
 Single pass sequencing.  
 This clone was obtained from a polyA-deleted cDNA library.  
**FEATURES**  
 source  
 1..563  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /clone="B1BR018A07"  
 /issue\_type="brain"  
 /dev\_stage="fetus"  
 /lab\_host="DH10B"  
 /clone\_lib="Bos taurus brain fetus"  
 /note="Vector: pZLI; Site\_1: SalI; Site\_2: NotI; Poly A was deleted from a NotI site"

**ORIGIN**  
 Query Match 4.2%; Score 65; DB 1; Length 563;  
 Best Local Similarity 56.2%; Pred. No. 0.00064;  
 Matches 122; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
 QY 533 GGACGAGTGTCCGCTGTGTGGAGCGGGAGCTATCTGTGTGACGGCTGCCCTCG 592  
 DB 28 GGAATATTGCGCGGTGTGCAAGAGCGGGAGAGCTGTGTGTGAGCGCTGCAATGTC 87  
 QY 593 GGGCTTCCACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652  
 DB 88 CTCCTACCAATCTCACTGTCTGAACCCCGCTGCCGACATCCCAACGGCAATGGCT 147  
 QY 653 GTGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 712  
 DB 148 GTGTCCCGCATGTGACCTGTCTGTCTCAAGAGCCGTGTGACAGAAATCTGTGACG 207  
 QY 713 GCCCCAGAGCCACCCGTGTGAGACCCGCTCCCCCG 749  
 DB 208 GTGGGAGAGACCCGCTGTGCGCATGCGACCCCGCCAG 244

**RESULT 21**  
**LOCUS** BE590137 610 bp mRNA linear EST 27-MAR-2003  
**DEFINITION** 197111 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.  
**ACCESSION** BE590137  
**VERSION** BE590137.1 GI:9843176  
**KEYWORDS** EST.  
**SOURCE** Bos taurus (cow)  
**ORGANISM** Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
**REFERENCE** 1 (bases 1 to 610)  
**AUTHORS** Sonstegard, T., Capuco, A.V., White, J., Van Tassell, C.P., Connor, E.E., Cho, J., Sultana, R., Shade, L., Wray, J.E., Wells, K.D. and Quakenbush, J.  
**TITLE** Analysis of bovine mammary gland EST and functional annotation of the Bos taurus gene index  
**JOURNAL** Mamm. Genome 13 (7), 373-379 (2002)  
**MEDLINE** 22135956  
**PUBMED** 12140684  
**COMMENT** Contact: Sonstegard TS  
 USDA, ARS, Beltsville Agricultural Research Center  
 Bldg. 200, Rm 2A, Beltsville, MD 20705, USA  
 Tel: 301 504 8416  
 Fax: 301 504 8414  
 Email: tads@lpsl.barc.usda.gov

**Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.**  
**PCR Primers**  
**FORWARD:** AGGAACAGCTATGACCAT  
**BACKWARD:** GTTTCCTGACGACG  
**Plate:** 117 row: I column: 4  
**Seq primer:** ATTAGGTGACCTATGAC.  
**FEATURES**  
 source  
 1..610  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /issue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="BARC 5BOV"  
 /note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI; library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

**ORIGIN**  
 Query Match 4.2%; Score 65; DB 2; Length 610;  
 Best Local Similarity 62.7%; Pred. No. 0.00064;  
 Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 532 AGGACAGTGTCCCTGTGTGTGGAGCGGGAGACTATCTGTGTGACGGCTGCCCTTC 591  
 DB 431 AGAATTAATCTGAGGTGTGCGACGAGGGGGAATATCTGTGCGACACTGCCCA 490  
 QY 592 GGGCTTCCACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651  
 DB 491 GGGCTTACATCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550  
 QY 652 GGTGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 692  
 DB 551 GCTGCCCCACATGTGAGAAGAGGAGATCAGTGGAGCGG 591

**RESULT 22**  
**LOCUS** CN322716/c 587 bp mRNA linear EST 14-APR-2004  
**DEFINITION** AGENCOURT 21860720 Xtst10-30 Xenopus tropicalis cDNA clone  
**ACCESSION** CN322716  
**VERSION** CN322716.1 GI:46380352  
**KEYWORDS** EST.  
**SOURCE** Xenopus tropicalis (western clawed frog)  
**ORGANISM** Xenopus tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae; Xenopodinae; Xenopus; Silurana.  
**REFERENCE** 1 (bases 1 to 587)  
**AUTHORS** NIH-MGC http://mgc.nci.nih.gov/.  
**TITLE** NIH-MGC  
**JOURNAL** National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
**COMMENT** Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cga@bbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Wei Wu / Prof. Christof NIEHRS  
 cDNA Library Preparation: Wei Wu, DKFZ  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:  
 http://image.llnl.gov  
 Plate: L1AM14930 row: k column: 22  
 High quality sequence stop: 587.  
**FEATURES**  
 source  
 1..587  
 /organism="Xenopus tropicalis"



```

Plate: ZF101-P00008-BR2 row: E column: 20
Seq primer: TTTT TTTT TTTT TTTT TTTT TTTT
High quality sequence stop: 807.
Location/Qualifiers
1. 807
source

```

**ORIGIN**

Query Match	4.1%;	Score 63.6;	DB 7;	Length 807;
Best Local Similarity	61.4%;	Pred. No. 0.0014;		
Matches 102; Conservative	0;	Mismatches 64;	Indels 0;	Gaps 0

QY	524	AAAGAAATGAGGACGAGTGTGCGGTGTGTGCGGACGGCGGGAGCATTGTGCTGTGACGG	583
Db	650	GGATTCATGAGATTTACTGTGAGGTCTGCGACGACGAGGAGGAGATCATTTCTTGTGACAC	591
QY	584	CTGCCCTCGGGGCTTCCACTGTGCGCTGTGCCCCCTCCGCTCGGGGAGATCCCAAGTGG	643
Db	590	CTGTCCACGGGGCTATTCACATGTGTGTGTGGACCCCGACATGAGAAAGGCCCTTGAAGGG	531
QY	644	GACCTGAGAGTGTCTCAGCTGCTGCAGGCAACAGTGCAGGAGGTG	689
Db	530	TACCTGGAGCTGCCACACTGTGAGAAATGTGGAAATCCAGTGGGAG	485

RESULT 25	AG030591	888 bp	DNA	linear	GSS 01-NOV-2001
LOCUS	AG030591				
DEFINITION	Pan troglodytes DNA, clone: PFB-003A10.F, genomic survey sequence.				
ACCESSION	AG030591				
VERSION	AG030591.1	GI:16557464			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes				
REFERENCE	Mumaiyola; Metaxaza; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalia; Eutheria; Primates; Catarrhini; Homidae; Pan.				
AUTHORS	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Matsunabe, H. and Sakaki, Y.				
TITLE	BAC end sequences of library PFB				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 888)				
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				

**TITLE** Totoki, Y., Watanabe, H. and Sakaki, Y.  
**Direct Submission**  
**JOURNAL** Submitted (02-AUG-2001) Aeao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-72 Shiehi-ryo-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [e-mail: chinbabe@gsc.riken.go.jp, URL: <http://hgp-gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170]  
**COMMENT** Clones are derived from the chimpanzee BAC library PBH This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.  
**PRIMERS**

**ORIGIN**

Query Match	4.1%	Score 63.6;	DB 9;	Length 888;
Best Local Similarity	44.2%;	Pred. No. 0.0014;		
Matches 269;	Conservative 0;	Mismatches 335;	Indels 4;	Gaps 1;

QY	693	CCCCGGGCAAGAGAGCCCGGCCCCAGAGGCAACCGGTGAGAGACCCGGCTCCCGGGG	752
Db	99	CCGGGGGCMCCCCGGGCCCCGGGCCCCCGGGCGCCCMCCGACCGGCGCGGGG	158
QY	753	CTTAGTGGGGGAGAGAGGTAAAGGTCACTTGGGGAACCCCTTAGCGGCATGGAC	812
Db	159	CCGGCCGCCCCCMNGGCGGGGTGGGGCCCCCCCCCGGCCCCCCCCCGCCGAA	218
QY	813	ACGACTTTGTCTTCAAGACACTGGGGGCTCGGCTTCTGAGGCCCCGCTGCAGGTCTG	872
Db	219	AAAAAAAAATCCTCCATACCMGCCCCGCCCCGGGCTGGCGCGCCCG---GGGNGCG	274
QY	873	GACTCCTCGGCGCTGCACCCCTTACTGTGTGGGTCTTGAAGGTCAAGACAACTGGCT	932
Db	275	GGGCMNNGGNCGCCGGGGCCCGGCCCCCGGCCCCCGGCGCGGCGCGGCGCCCCC	334
QY	933	CTGTGTGCGGTTGCGGGGTGTGTGCGAGATGTATCGAAGTGTGCGGTATCTCATCTGC	992
Db	335	CCCCGGCGGGCGCGCGCGCGCGCCGCMNGGGCGAGNGCGAGGCGGNCGCCCGCCG	394
QY	993	GGCGCTGCTTTCACATGAGCGCTGCATCTCCAGCGGAGCACTCCGGGCGGGAGGAGC	1053
Db	395	CCCGCGGCGCCCCCGGCCCCCGGCCCCCGGCGCCGCMNCCCCCGCGGCGGGCCCCG	454
QY	1053	CTGGCTTGCAGATCTTCTCTGAGAGAGTGAACCCAGGCGCTGTGAGAGGGGTGTGAGC	1113
Db	455	CCCCGCCCCCCCCCGGGGCCCCCCCCCGGCCCCCGGCGCCCCCGCGCGCG	514
QY	1113	CCCAAGCCCCCGCGGCTTGGCTCTTGGGGCTTGCAGAGATGACATGCGCATGTCAGAGCC	1173
Db	515	GCCCCCCCCCGGCGGCCCCCGGCCCCCGGCGCCCGCMNCCCCCGGCGCGCCCCCGCGCG	574
QY	1173	GCTTGTGACAGAGATGACTTGAAGTCCCTTTTGAAGGAGACACCTTGATGGACATCTG	1233
Db	575	CCGGCGCGGAGCCCCCGGCCCCCGGCGCGCGCGCGCCCCCGGCGCGCGCGCGCG	634
QY	1233	CAGTGGGCAATCCAGAGCATGGCCGTCGCGGCGCCCCCTTCCCTCTGTGACCCCAAGTG	1293
Db	635	GGGCGGCGCGGCGCGCGCGCGCCCCCGGCGCGCCCCCGGCCCCCGGCCCCCGG	694
QY	1293	GGCGGGAGC 1300	

Db 695 GCGGCGCC 702

RESULT 26  
LOCUS BX915120  
DEFINITION 783 bp mRNA linear EST 07-MAY-2004  
BX915120 Sus scrofa library (scan) Sus scrofa cDNA clone  
scan0028d.a.12 5prim, mRNA sequence.

ACCESSION BX915120  
VERSION BX915120.1 GI:41131899  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa (pig)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
REFERENCE 1 (bases 1 to 783)  
Bonnert, R., Tosser-Klopp, G., Benne, F., Cabau, C., Villegier, S.,  
Soares, M., Bonaldo, F. and Harey, F.  
A Pig Normalised Multi-Tissue cDNA Library  
Unpublished (2003)  
CONTACT: Tosser-Klopp G  
Genetique Animale  
Institut National de la Recherche Agronomique  
Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan  
cedex, FRANCE  
Tel: 33 (0) 5.61.28.51.14  
Fax: 33 (0) 5.61.28.53.08  
Email: tosser@toulouse.inra.fr  
Sequence cleaned of vector, adaptor and repetitions. Contact us  
at sigenauport@jouy.inra.fr to obtain the chromatogram of this  
sequence.  
Plate: 0028 row: 8 column: 12.  
Location/Qualifiers  
1. .783  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone="scan0028d.a.12"  
/tissue\_type="mixed"  
/dev\_stage="from embryos to adults"  
/clone\_lib="Sus scrofa library (scan)"  
/note="Tissues: adipose tissue, brain, kidney, liver,  
muscle, ovary, testis, heart, hypothalamus, pancreas,  
skin, spleen, thymus, placenta, pituitary gland, seminal  
vesicle, small intestine, uterus, adrenals, bulbourethral  
gland, cerebral trunk, epididymis, female gonad,  
gall-bladder, hippocampus, large intestine, male gonad,  
melanocytes, stomach, udder"

ORIGIN

Query Match 4.1%; Score 63.4; DB 5; Length 783;  
Best Local Similarity 61.7%; Pred. No. 0.0015;  
Matches 100; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 528 AATGAGAGAGGTCCTGCTGTCGGAGCGGGAGAGCTATGCTGTGACGCTCG 587  
|||||  
Db 585 AATGAGAGCTGCTGCTGCTGTCAGACGAGGGAATCTGCTGTGGAAGTG 644  
|||||  
QY 588 CCTCGGGCCTTCCACCTGGCTGCTGCTCCCTCCGCTCGGAGATCCCACTGGAGCC 647  
|||||  
Db 645 CCGAAGTATTCACCTGCTGTCACGTCGCCACCTGGCAATTTTCCGAGTGGAGG 704  
|||||  
QY 648 TGGAGGTGCTCCAGCTGCTGCAAGCAAGTCAGAGGTG 689  
|||||  
Db 705 TGGATTGACATTTCTGCGAGACTTATCCNAACAGAAAGTG 746  
|||||

RESULT 27  
LOCUS AA792266  
DEFINITION 406 bp mRNA linear EST 09-FEB-1998  
vn94e11.r1 StrataGene mouse heart (#937316) Mus musculus cDNA clone  
IMAGE:1039628 5', mRNA sequence.  
ACCESSION AA792266

VERSION AA792266.1 GI:2855221  
EST.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 406)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
CONTACT: Marras M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:579652  
Seq primer: -28m3 rev1 ET from Amersham  
High quality sequence stop: 403.  
Location/Qualifiers  
1. .406  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NIH Swiss"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1039628"  
/sex="pooled"  
/tissue\_type="heart"  
/dev\_stage="13 day embryos"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse heart (#937316)"  
/note="Organ: heart; Vector: pBluescript SK-; Site 1:  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dr. 93 pooled NIH/Swiss 13 day embryo hearts.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'  
adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

ORIGIN

Query Match 4.1%; Score 63.2; DB 1; Length 406;  
Best Local Similarity 67.4%; Pred. No. 0.0016;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGCAGGTGCTGCTGTCGGAGCGGGAGAGCTATGCTGTGACGCTGCCCTCG 592  
|||||  
Db 220 GGAATCTCTCCGCTGTCGAGAGCGGGAGAGCTCTGCTGTGACATGCGCTTC 279  
|||||  
QY 593 GGCCTTCCACCTGCTGCTGCTCCCTCCGCTCGGAGATCCCACTGGAGCC 652  
|||||  
Db 280 TTCTTACCACTTCACTGCTGAACCCCGCTGCGAGATCCCAACGGGGAATGCT 339  
|||||  
QY 653 GTGCTTCACCTG 664  
|||||  
Db 340 CTGTCCCGCTG 351  
|||||

RESULT 28  
LOCUS CF161888  
DEFINITION 485 bp mRNA linear EST 25-JUL-2003  
B0703B06-5 NIA Mouse Embryonic Germ Cell cDNA library (long) Mus  
musculus cDNA clone NIA:B0703B06 IMAGE:30458165 5', mRNA sequence.  
ACCESSION CF161888  
VERSION CF161888.1 GI:33271437  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 485)  
Piao, Y., Ko, N.-T., Lim, M. K. and Ko, M. S. H.  
Construction of long-transcript enriched cDNA libraries from  
submicrogram amounts of total RNAs by a universal PCR amplification  
method

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Genome Res. 11 (9), 1553-1558 (2001)  
21429098  
11541199  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: B0703 row: E column: 06  
Seq primer: M13 Reverse  
High quality sequence stop: 485  
POLYA=No.

FEATURES  
source

1. 485  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="taxon:10090"  
/clone="NIA:B0703B06 IMAGE:30458165"  
/sex="male"  
/dev\_stage="embryonic day 8"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Embryonic Germ Cell cDNA library  
(long)"  
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).  
This is a long-transcript enriched cDNA library (Ref.  
Genome Res. 11: 1553-1558 (2001). [PMID: 11541199]). Total  
RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP).  
EG cells were cultured at 37. C, 5% CO2 in DMEM  
supplemented with 15% ES cell-qualified FBS, 0.1mM  
non-essential amino acids, 2 mM glutamine,  
penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM  
beta-mercaptoethanol, and 1000000 units of LIF per liter.  
Double-stranded cDNAs were synthesized with an Oligo(dT)  
primer (Invitrogen):  
5'-pGACTGATCTTGAATCGCAGCGCGCCCTTTT-3' from  
2.5 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to lona-linker LL-SalI, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer SalI-4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 4.0 kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 485;  
Best Local Similarity 67.4%; Pred. No. 0.0016;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGGTGCTGCTGTCGAGCGCGGAGCTATCTGTCGAGCGCTCGC 592  
DB 55 GGAATCTGTGCGCTGTCGAGCGCGGAGCTCTGCTGTGACACATGCCCTTC 114  
QY 593 GGCCTTCACCTGAGCTGCTGTCCTCCCTCGCTCCGAGAGATCCCACTGAG 652  
DB 115 TTCTACACATCCATGCTGAAACCCCGCTGCCAGATCCAAACGCGAATGGCT 174

QY 653 GGTCTCCAGCTG 664  
DB 175 GTGTCCCGCTG 186

RESULT 29  
CA893864  
LOCUS  
DEFINITION

CA893864 514 bp mRNA linear EST 20-DEC-2002  
B0181H02-5N NIA Mouse Neural Stem Cell (Differentiated) cDNA  
Library (long) Mus musculus cDNA clone NIA:B0181H02 IMAGE:30101653  
5', mRNA sequence.

ACCESSION  
CA893864  
VERSION  
CA893864.1 GI:27345413  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 514)  
Piao, Y., Dudekula, D. B., Qian, Y., Martin, P. R., Aiba, K., Vescevi, A. L.  
and Ko, M. S. H.  
Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)  
cDNA Library (long)  
Unpublished (2002)

JOURNAL  
COMMENT

Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: B0181 row: H column: 02  
Seq primer: -21M13 Reverse  
High quality sequence stop: 514  
POLYA=No.

FEATURES  
source

1. 514  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CD1"  
/db\_xref="taxon:10090"  
/db\_xref="taxon:10090"  
/clone="NIA:B0181H02 IMAGE:30101653"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Neural Stem Cell (Differentiated)  
cDNA Library (long)"  
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:  
NotI; Mouse cDNA project by the Laboratory of Genetics,  
National Institute on Aging (NIA), Intramural Research  
Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is  
a long-transcript enriched cDNA library (Ref. Genome Res.  
11: 1553-1558 (2001). [PMID: 11541199]). Total RNAs were  
obtained from Dr. Angelo L. Vescevi (Institute for Stem  
Cell Research, Italy). Double-stranded cDNAs were  
synthesized with an Oligo(dT) primer (Invitrogen):  
5'-pGACTGATCTTGAATCGCAGCGCGCCCTTTT-3' from  
2.0 microgram of total RNA, treated with T4 DNA  
polymerase, and purified by ethanol-precipitation. The  
cDNAs were ligated to lona-linker LL-SalI, purified by  
phenol/chloroform, and separated from free linkers by  
Centricon 100. Then, the cDNAs were amplified by  
long-range high fidelity PCR using Ex Taq polymerase  
(Takara) with a primer SalI-4-S. The products were purified  
by phenol/chloroform and Centricon 100. The cDNAs were  
digested with SalI and NotI enzymes and cloned into  
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.  
coli host was transformed with the ligation mixture by the  
standard chemical method. The average insert size is about  
3.2 kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 514;  
Best Local Similarity 67.4%; Pred. No. 0.0016;



	Matches	89;	Conservative	0;	Mismatches	43;	Indels	0;	Gaps	0;
OY	533	GGACGAGTGTGCCCTGTGTGCGGAGACGGGGGAGCTCATTCGTCTGTACAGGCTGCCCTCG	592							
Db	312	GGAAATTCTGTGTGCGCTGTCCAAAGSAGGGGGGAGCTCTCTGTGCTGTGACACATGCGCCCTTC	371							
OY	593	GACCTTTCCACCTGGGCTGCTGCTGCCCTCCGCTCCGGAGAGATCCCAATGGAGACTGTGAG	652							
Db	372	TTTCTTACACATTCACCTGCTGTGAACCCCCCGCTGTCCAGATGATCCCAACGGGGAATGGCT	431							
OY	653	GTGCTTCAGCTG	664							
Db	432	CTGTCCCGCTG	443							

RESULT 30					
CAS64994					
LOCUS	CAS64994	562 bp	mRNA	linear	EST 19-NOV-2002
DEFINITION	K0332G02-5N NIA Mouse Osteoblast cDNA library (Long) Mus musculus				
	cDNA clone NIA:K0332G02 IMAGE:30057385 5', mRNA sequence.				
ACCESSION	CAC64004				

VERSION	CAS64994.1	GI:25109673
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS  
1 (bases 1 to 562)  
Piao, Y., Kargul, G. J., Dudekula, D. B., Qian, Y., Luo, A., Carter, M. G.

TITLE	Systematic Analyses of NIA Mouse Osteoblast cDNA Library (Long
JOURNAL	Unpublished (2001)
COMMENT	Other_ESTs: K0332G02-3

Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdn@n1gsun.gsc.nia.nih.gov  
Plate: K0332 Row: G Column: 02  
Seq primer: M13 Reverse  
High quality sequence stop: 562  
POLYA=NO.

FEATURES	Location/Qualifiers
source	1. .562

```

1. 562
/organism="Mus musculus"
/mol_type="mrna"
/strain="C3H/He"
/db_xref="naest:K033G02-5N"
/db_xref="taxon:10090"
/clone="NIA:K033G02 IMAGE:30057385"
/tissue_type="osteoblast"
/cell_line="KUSA-A1 cells"
/lab_host="DH10B"
/clone_1fb="NIA Mouse Osteoblast cDNA Library (Long)"
/notes="Vector: pSPORT (Invitrogen); Site 1: SalI; Site 2:
NotI. Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://ligsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). (PMID: 11544199)). Total RNAs were
obtained from Dr. Akihiro Umezawa (Keio University School
of Medicine, Japan). Double-stranded cDNAs were
synthesized with an oligo(dT) primer (Invitrogen:
5'-pGACAGTGTGATCGATCGGACGCCGCCCTTTTCTTTT-3') from
2.1 µg of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker UL-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-5. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT plasmid vector.

```

**ORIGIN**

Query Match	4.1%;	Score.63.2;	DB 6;	length 562;
Best Local Similarity	67.4%;	Pred. No. 0.0016;		
Matches	89;	Conservative	0;	Mismatches 43;
			Indels	0;
			Gaps	0;

Qy	533	GGACGAGTATGCGCGTGTGTGCGGAGCGGGGGGAGGTCAATCTGCTGTGACGGCTCCCTCG	532
Db	73	GGAAATCTCTCCGCTGTGCAGAGAGGGGGGAGCTCTGTGCTGTGACATCCCTTC	132
Qy	593	GGACCTTCACACTGACTGCTGTCCCTTCGCTCCGGAGATCCCGCAGTGGAGCTTGAG	652
Db	133	TTCCTACCAACATCCACTGCTGAAACCCCGCTGCAAGATCCCAACGGCGAATGCT	132
Qy	653	GTGCTCCACTG	664
Db	193	CTGTCCCGCTG	204

RESULT 31  
BB654405

LOCUS	BB654405	564 bp	mRNA	linear	EST 26-OCT-2001
DEFINITION	BB654405	Riken full-length enriched, 2 days neonate thymus thymic cells Mus musculus cDNA clone G920008E20 5', mRNA sequence.			
ACCESSION	BB654405				

VERSION	BB654405.1	GI:16488233
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

## REFERENCE AUTHORS

TITLE	JOURNAL	COMMENT
Unpublished (2001)		
Contact: Yoshihide Hayaehizaki		
RIKEN Mouse ESTs (Araiaka, T., et al. 2001)		
		<p> Hara, A., Hiramoto, K., Horii, F., Ihii, Y., Ito, M., Kawai, J.,  Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  Ohno, H., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  Sano, H., Sasaki, D., Shibata, K., Shingana, A., Shiraki, T.,  Takada, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  Takeda, Y., Tanaka, T., Toyoi, T., Muramatsu, M., and Hayaehizaki, Y. </p>

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
wagi, K., Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
Kondo, S., Shitagawa, A., Saito, T., Kiyosawa, H., Yamataka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin Unive  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouse1.html>  
This clone was contributed by the Brain Molecular Anatomy Project

FEATURES  
source

(BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers

1. .625  
/organism="Mus musculus"  
/mol\_type="rRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30630778"  
/issue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_11b="NIH BMAP\_HB0"  
/note="Organ: Eye; Vector: PYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTTATGAGCT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 7; Length 625;  
Best Local Similarity 67.4%; Pred. No. 0.0016; Mismatches 43; Indels 0; Gaps 0;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTGCGGTGTGTCGGAGCGCGGAGCTCATCTGTGACGCGCTCG 592  
|||  
DB 179 GGAATCTCTCGGTGTGCAAGAGCGCGGAGCTCCTGTGTGACATGCCCTTC 238  
|||  
QY 593 GGCCTTCCACCTGCGCTGCTCCCTCCCTCCGCGGAGATCCCAATGGAGCTGAG 652  
|||  
DB 239 TTCTCAACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGGATGGCT 298  
|||  
QY 653 GTGCTCCAGCTG 664  
|||  
DB 299 CTGTCCCGCTG 310

RESULT 34  
BU05998 634 bp mRNA linear EST 15-JUL-2003  
LOCUS BU05998  
DEFINITION UI-M-F00-caa-1-10-0-UI.r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
IMAGE:6408657 5', mRNA sequence.  
ACCESSION BU05998  
VERSION BU05998.1 GI:23636034  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 634)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source

(BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers

1. .634  
/organism="Mus musculus"  
/mol\_type="rRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6408657"  
/issue\_type="whole brain"  
/dev\_stage="embryo 12.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_11b="NIH BMAP\_F00"  
/note="Organ: Brain; Vector: PYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAAGAGCC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 5; Length 634;  
Best Local Similarity 67.4%; Pred. No. 0.0016; Mismatches 43; Indels 0; Gaps 0;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTGCGGTGTGTCGGAGCGCGGAGCTCATCTGTGACGCGCTCG 592  
|||  
DB 269 GGAATCTCTCGGTGTGCAAGAGCGCGGAGCTCCTGTGTGACATGCCCTTC 328  
|||  
QY 593 GGCCTTCCACCTGCGCTGCTCCCTCCCTCCGCGGAGATCCCAATGGAGCTGAG 652  
|||  
DB 329 TTCTCAACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGGATGGCT 388  
|||  
QY 653 GTGCTCCAGCTG 664  
|||  
DB 389 CTGTCCCGCTG 400

RESULT 35  
BU055310 635 bp mRNA linear EST 26-AUG-2002  
LOCUS BU055310  
DEFINITION UI-M-F00-bzp-b-16-0-UI.r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
IMAGE:6405423 5', mRNA sequence.  
ACCESSION BU055310  
VERSION BU055310.1 GI:22495387  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 635)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES Seq primer: pyx-5.  
Location/Qualifiers  
source 1..635  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6405423"  
/issue\_type="whole brain"  
/dev\_stage="embryo 12.5dpc"  
/lab\_host="NIH BMAP F00"  
/note="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TAGAGAGCC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 5; Length 635;  
Best Local Similarity 67.4%; Pred. No. 0.0016;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 533 GGACGAGTGGCCGTGTGTCGGAGCGCGGAGCTCATCTGCTGAGCGGCTGCGCTCG 592  
DB 14 GGAAATTCGTCCGCTGTGAGAGCGCGGAGCTCTGCTGTGACACATGCGCTTC 73  
QY 593 GGCTTCACACCTGGCTGTGCTGCTCCCTCGCTCGGAGATCCCGAGTGGAGCTGAG 652  
DB 74 TTCTTACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCAAAGCGGAATGGCT 133  
QY 653 GTGCTCCAGCTG 664  
DB 134 CTGTCCCGCTG 145

RESULT 36  
CB246896 723 bp mRNA linear EST 09-JUL-2003  
LOCUS UI-M-FIO-cdy-K-21-0-UI.r1 NIH BMAP\_FIO Mus musculus cDNA clone  
DEFINITION IMAGE:6836206 5', mRNA sequence.  
ACCESSION CB246896  
VERSION CB246896.1 GI:28368540  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 723)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES Seq primer: pyx-5.  
Location/Qualifiers  
source 1..723  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6836206"  
/issue\_type="whole brain"  
/dev\_stage="embryo 12.5dpc"  
/lab\_host="NIH BMAP F10"  
/note="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCAGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 723;  
Best Local Similarity 67.4%; Pred. No. 0.0017;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 533 GGACGAGTGGCCGTGTGTCGGAGCGCGGAGCTCATCTGCTGAGCGGCTGCGCTCG 592  
DB 357 GGAAATTCGTCCGCTGTGAGAGCGCGGAGCTCTGCTGTGACACATGCGCTTC 416  
QY 593 GGCTTCACACCTGGCTGTGCTGCTCCCTCGCTCGGAGATCCCGAGTGGAGCTGAG 652  
DB 417 TTCTTACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCAAAGCGGAATGGCT 476  
QY 653 GTGCTCCAGCTG 664  
DB 477 CTGTCCCGCTG 488

RESULT 37  
CF726532 730 bp mRNA linear EST 09-OCT-2003  
LOCUS UI-M-HB0-cdh-h-09-0-UI.r1 NIH BMAP\_HB0 Mus musculus cDNA clone  
DEFINITION IMAGE:30547664 5', mRNA sequence.  
ACCESSION CF726532  
VERSION CF726532.1 GI:37600700  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 730)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyx-5.  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source

Location/Qualifiers  
1..730  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30547664"  
/issue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1ib="NIH BMAP H50"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTAATGAGT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 7; Length 730;  
Best Local Similarity 67.4%; Pred. No. 0.0017;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTCCGCTGTGTCGGAGCGCGGAGCTCATCTGTGTGACGGCTGCCCTGG 592  
DB 51 GGAAATTCCTCCGCTGTGACAGACGCGGAGCTCTGTGTGTGACACATGCCCTTC 110  
QY 593 GGCGTTCCACCTGCGCTGCTGCCCTCCGCTCCGAGGATCCGAGTGGAGCTGGAG 652  
DB 111 TTCTTACCACTCATCTGCTGTAACCCCGCTGCCAGAGATCCCAACGCGCAATGGCT 170  
QY 653 GTGCTCCAGCTG 664  
DB 171 CTGTCCCGCGCTG 182

RESULT 38  
LOCUS CN534827 736 bp mRNA linear EST 29-APR-2004  
DEFINITION UI-M-HSO-cgd-k-04-0-UI.r1 NIH\_BMAP\_HSO Mus musculus cDNA clone  
IMAGE:30674835 5', mRNA sequence.  
ACCESSION CN534827  
VERSION CN534827.1 GI:46862983  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 736)  
TITLE NIH-MGC http://mgc.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mouseefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
1..736  
Location/Qualifiers

## ORIGIN

Query Match 4.1%; Score 63.2; DB 7; Length 736;  
Best Local Similarity 67.4%; Pred. No. 0.0017;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTCCGCTGTGTCGGAGCGCGGAGCTCATCTGTGTGACGGCTGCCCTGG 592  
DB 46 GGAAATTCCTCCGCTGTGACAGACGCGGAGCTCTGTGTGTGACACATGCCCTTC 105  
QY 593 GGCGTTCCACCTGCGCTGCTGCCCTCCGCTCCGAGGATCCGAGTGGAGCTGGAG 652  
DB 106 TTCTTACCACTCATCTGCTGTAACCCCGCTGCCAGAGATCCCAACGCGCAATGGCT 165  
QY 653 GTGCTCCAGCTG 664  
DB 166 CTGTCCCGCGCTG 177

RESULT 39  
LOCUS CO045219 754 bp mRNA linear EST 10-JUN-2004  
DEFINITION UI-M-HO0-cmx-0-19-0-UI.r1 NIH\_BMAP\_HO0 Mus musculus cDNA clone  
IMAGE:30638082 5', mRNA sequence.  
ACCESSION CO045219  
VERSION CO045219.1 GI:46585373  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 754)  
TITLE NIH-MGC http://mgc.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mouseefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
1..754  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"

/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30638082"  
/tissue\_type="Upper Head"  
/dev\_stage="9.5-10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1ib="NIH BMAP H00"  
/note="Organ: Head; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 7; Length 754;  
Best Local Similarity 67.4%; Pred. No. 0.0017;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGGAGGATCATCTGTGACGCGCTCG 592  
DB 39 GGAATTTCTTCCGCTGTGCAAGAGCGCGGAGCTCTGTGTGACACATGCCCTTC 98  
QY 593 GGCCTTCCACCTGAGCTGTGCTCCCTCGCTCCGGAGATCCCGATGGAGCTGAG 652  
DB 99 TTCTTACCACTGCACTGCTGAACCCCGCTGCGAGATCCAAAGCGGAATGGCT 158  
QY 653 GTGCTCCAGCTG 664  
DB 159 CTGTCCCGCGCTG 170

RESULT 40  
LOCUS CN536888 755 bp mRNA linear EST 29-APR-2004  
DEFINITION UI-M-HO0-cqb-3-11-0-UI.r1 NIH BMAP\_HO0 Mus musculus cDNA clone  
IMAGE:30664450 5', mRNA sequence.  
ACCESSION CN536888  
VERSION CN536888.1 GI:46865044  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
AUTHORS 1 (bases 1 to 755)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabp-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
Location/Qualifiers  
1..755  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"

/clone="IMAGE:30664450"  
/tissue\_type="Upper Head"  
/dev\_stage="9.5-10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1ib="NIH BMAP H00"  
/note="Organ: Head; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 7; Length 755;  
Best Local Similarity 67.4%; Pred. No. 0.0017;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGGAGGATCATCTGTGACGCGCTCG 592  
DB 46 GGAATTTCTTCCGCTGTGCAAGAGCGCGGAGCTCTGTGTGACACATGCCCTTC 105  
QY 593 GGCCTTCCACCTGAGCTGTGCTCCCTCGCTCCGGAGATCCCGATGGAGCTGAG 652  
DB 106 TTCTTACCACTGCACTGCTGAACCCCGCTGCGAGATCCAAAGCGGAATGGCT 165  
QY 653 GTGCTCCAGCTG 664  
DB 166 CTGTCCCGCGCTG 177

RESULT 41  
LOCUS CA327845 762 bp mRNA linear EST 09-UTL-2003  
DEFINITION UI-M-FY0-ccw-f-03-0-UI.r1 NIH BMAP\_FY0 Mus musculus cDNA clone  
IMAGE:6825316 5', mRNA sequence.  
ACCESSION CA327845  
VERSION CA327845.1 GI:24545943  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.  
AUTHORS 1 (bases 1 to 762)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabp-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source  
Location/Qualifiers  
1..762  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6825316"

/cissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP FY0"  
/note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonafio, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction. ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Aac vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AGCGAGACAG. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,  
program coordinator."

## ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 762;  
Best Local Similarity 67.4%; Pred. No. 0.0017;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGAGTGTCCGCTGTGCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTCG 552  
DB 342 GGAATTCCTGTGCGCTGTGCAAGGAGCGCGGAGCTCTGTGCTGTGACACATGCCCTTC 401  
QY 593 GGGCTTCACCTGCGCTGTGCTGCCCTCGCTCCGCGGAGATCCCAAGTGGAGACTGGAG 652  
DB 402 TTCTTACCACTACCTGCTGAACCCCGCTGCCAGATCCCAAGCGGCAATGGCT 461  
QY 653 GTGCTCCAGCTG 664  
DB 462 CTGTCCCGCTG 473

RESULT 42  
CA749488 768 bp mRNA linear EST 09-JUL-2003  
LOCUS UI-M-FYO-cdd-n-01-0-UI.r1 NIH\_BMAP\_FYO Mus musculus cDNA clone  
DEFINITION IMAGE:6831650 5', mRNA sequence.  
ACCESSION CA749488.1 GI:25571583  
VERSION CA749488.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS NIH-MGC  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
http://image.lnll.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.  
Location/Qualifiers  
1..768  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6831650"

## FEATURES

source

## ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 768;  
Best Local Similarity 67.4%; Pred. No. 0.0017;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGAGTGTCCGCTGTGCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTCG 552  
DB 97 GGAATTCCTGTGCGCTGTGCAAGGAGCGCGGAGCTCTGTGCTGTGACACATGCCCTTC 156  
QY 593 GGGCTTCACCTGCGCTGTGCTGCCCTCGCTCCGCGGAGATCCCAAGTGGAGACTGGAG 652  
DB 157 TTCTTACCACTACCTGCTGAACCCCGCTGCCAGATCCCAAGCGGCAATGGCT 216  
QY 653 GTGCTCCAGCTG 664  
DB 217 CTGTCCCGCTG 228

RESULT 43  
CK637900 787 bp mRNA linear EST 28-JAN-2004  
LOCUS UI-M-HOO-cnt-m-12-0-UI.r1 NIH\_BMAP\_HOO Mus musculus cDNA clone  
DEFINITION IMAGE:30642251 5', mRNA sequence.  
ACCESSION CK637900.1 GI:41363766  
VERSION CK637900.1  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS NIH-MGC  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mouse1.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.  
Location/Qualifiers  
1..787  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30642251"  
/cissue\_type="Upper Head"

## FEATURES

source







Db 529 MYSSSVSCSCSGCTCCGYCSCTSCMKSCCTGCKGCGCGCTSCSSSSCCSBBSYST 588  
 Qy 935 TGGTGGCGCTTGGGGGTGGGAGATGGTACGATGCTGCGGTGATCTACATGCGC 994  
 Db 589 CCGTCTCTKCSGGCTSGCTGCGGGGCGCGGGGCGGGGCGGGGCGGGGCGGGGCGG 648  
 Qy 995 CGCTGCTTCCACTGAGCGCTGACCTTCCAGCCGCGACTCCCGGCGCGAGACGGAGCCT 1054  
 Db 649 SSGSSSSSSCGSSSSSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 708  
 Qy 1055 GCGCTGCAATCTCTGCTCAGAGACGTGATCCAGCCCTCTGTGTGAGAGGGGTCTGCGCCC 1114  
 Db 709 SGGGCGSCGSCSGCGSSCGSSSSSGCCSCGSCSSSSSSCGSCCGCGGMSGSG 768  
 Qy 1115 CAGCCCGCGCGCTGCGCCCTGCGCTGCGCTGCGAGATGACATGACATGACGAGCGCGC 1174  
 Db 769 CGSSSGCG 828  
 Qy 1175 TCTGCAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1234  
 Db 829 CGGGCG 888  
 Qy 1235 GTGGCGCATTCAGAGATGAGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1280  
 Db 889 GCGSGGSGGSGCGSGGSGGSGCGSGCGSGGSGGSGGSGGSGGSGGSGGSGGSGG 934

RESULT 48  
 CR715927 1066 bp mRNA linear HTC 12-AUG-2004  
 LOCUS CR715927  
 DEFINITION Tetraodon nigroviridis full-length cDNA.  
 ACCESSION CR715927.1 GI:51214161  
 VERSION CR715927.1  
 KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.  
 SOURCE Tetraodon  
 ORGANISM Tetraodon  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae.  
 1 (bases 1 to 1066)  
 Genoscope.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT The sequences are based on single pass reads.  
 More information available at  
 http://www.genoscope.cns.fr/tetraodon.  
 FEATURES  
 source  
 1.1066  
 /organism="Tetraodon"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:47144"  
 /tissue\_type="Eggs"

ORIGIN  
 Query Match 4.1%; Score 63.2; DB 3; Length 1066;  
 Best Local Similarity 56.9%; Pred. No. 0.0017;  
 Matches 116; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 546 GTGTGTGGGAGCGCGGGAGCTCATCTGCTGTCGAGCGGCTTCGAGCTTCACCTG 605  
 Db 229 GTGTGCAAGATGAGAGAGAGCTTGTGCTGTGACACCTGCCCCCTCTCTACACATC 288  
 Qy 606 GCGTGTCTGCTCCCTCGCTCCGGAGATCCCAAGTGGAGCTGTCCAGCTGC 665  
 Db 289 CACTGCTCAACCTCCCTCCCTCCGAATCCCTATGAGAAATGATCTGCCCCCGCTGC 348  
 Qy 666 CTGACGCAACAGTCCGAGAGTGCACACCCCGGCAAGAGACCCCGCCCGAGAGCCA 725  
 Db 349 AAGGTCTCACCAATGAGAGGGAAGTTTCAAGAAATTTTAACTTGGCGATGGGGGACCA 408

Qy 726 CCCGTGAGACCCCGCTCCCCCG 749  
 Db 409 CCCGCCCCACACTGTCTCTCG 432

RESULT 49  
 AK034549 3137 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK034549  
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA. RIKEN full-length enriched library, clone:943004K15 product:CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2 AUTOMANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog [Homo sapiens], full insert sequence.  
 AK034549  
 VERSION AK034549.1 GI:26330020  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20493374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Komano, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861

4  
 THE RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE  
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 REFERENCE  
 AUTHORS Nature 420, 563-573 (2002)  
 5 (bases 1 to 3137)  
 Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komano, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://famom.gsc.riken.jp/>

## FEATURES

## source

Location/Qualifiers  
1..3137  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:9430004K15"  
/db\_xref="taxon:10090"  
/clone="9430004K15"  
/issue\_type="embryonic body between diaphragm region and neck"  
/clone\_1b="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
1..3137  
/note="unlabeled protein product; CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2 AUTOANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog [Homo sapiens] (SWISSPROT|Q14839, evidence: PASTY, 99.8%ID, 54.6%length, match=3135)  
putative"  
/codon\_start=3  
/protein\_id="BAC28749.1"  
/db\_xref="GI:26330021"  
/translation="PIRKAKTEGKGPNNRRKSGSPVPADAKPKPKVAPKIKLG  
GSGSRKSSSEDDLDVSDPDDASINSTSVDSSTSRSRKRLTKRKKEE  
EYAVADVETDHDQCEVCOQGEIILCDPCPAYMVCIDPMKAPKSCPCSE  
KEGIQWEAKEDNESEGEILEVGEDEEDHMEFCVCKDGEILCCDTPSSYH  
HCLNPLPEIPIENBMLCPKCTCPALGKQKILIMKGGPSPTPVPRPDNPNS  
PKLEBGRPEROPFVKMGQSYHMSVSELRLHCOMPRNRYORNDMDPEPSG  
GDEBKRKRNKDPKPAEMBERPYRGICPEMMATIRILNHSYDKGHYILKMDL  
PDAQWESBDEVIDYDLEFKOSYMHRELMESEBPGKILKVKILKLEPPEITP  
VPTVYERQPELTATGTGLHPYQWEGMLWLFSAQGTDTILADMGKXPTAV  
FVSLYKESKSGPFLVSAPLSTIIMEREFEWADMDVYTVVGDGSRALIREHF  
SFDNALRGKSKRMKEASVKEFVILTSYELITDMAIIGSIDMACILVDEARLK  
NNOSKPRVLNGYSLOHKLILGTPLONNLELFIHLPLTPPERFNLISGLEPADI  
AKEDQIKKLDHNLGPHMLRLKADVFRKNPSKTEILVRYELSMOKYIYITRPE  
ALVARGGQVQSVLNVVMDKCCNHPYLFVVAEMAPKPNMAYGSAITIRSGKL  
LLOKMLKNLEGGHRLVIFSQMTKMLDLDELPLEHGYKTERIDGQITGMREARLDR  
FNAPOQCFPLSTRAIGIATADTVIIVDSQWPNNDIQASRAIRIQONKRV  
MYRFTVTRASVEERTIOVAKKQMLTHLVRLPGLSGTSGMSQDLIDILKFTSELF  
KDEATDGGDNKGEDESVLHYDKAIERLDNLQDETDTLQGNVYLSKFAVQY  
VREEMGESEEVER"

## CDS

Query Match 4.1%; Score 63.2; DB 3; Length 3137;  
Best Local Similarity 67.4%; Pred. No. 0.002;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

## ORIGIN

533 GAGCAGGTGTCCTGTGTGGGACGGCGGGAGCTCATCTGCTGACGGCTGCGCTCG 592  
584 GGAATCTGTCTGTCTGCAAGGACGGCGGGAGCTCTCTGTGTCGACATGACCTTC 643  
593 GGCCTTCCACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652  
644 TTCTTACCATCATCTGCTGAAACCCCGCTGCGACAGATCCCAACGGCAATGGCT 703  
653 GTGCTCCAGCTG 664  
704 CTGTCCCGCTG 715

RESULT 50  
CNS0091P/c 925 bp DNA linear GSS 03-JUN-1999  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19016 of RPCT-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
AL053013  
VERSION  
AL053013.1 GI:4934461  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)

REFERENCE  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

## COMMENT

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osodegawa and  
Aaron Mammosier in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCT-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## source

Location/Qualifiers  
1..925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19016"  
/clone\_1b="RPCT-98"  
/note="end : TET3"

## ORIGIN

Query Match 4.1%; Score 62.8; DB 9; Length 925;  
Best Local Similarity 13.3%; Pred. No. 0.0021;  
Matches 51; Conservative 181; Mismatches 152; Indels 0; Gaps 0;

436 GCTGGGGAGCTGTTTGGGAAGAGGTGCTCTCAGAGGTGCTGACCCAGCCAG 495  
916 SCSSBSCSSSSMTSSNSBSCSSSSSSSTSSMSSSSSSSSSSSSSGTSACV 857  
496 TCTGCATGGCGCTCTTCTCTGCTGCAAGAATGAGACAGAGTTCCTGTGCGG 555  
856 KCNASSSCCGCGGABCCMCSSSSCCGSAARGVKVAAGAGRRGGSGASASHS 797  
556 ACCGCGGAGAGCTCATCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 615  
796 SNAACSSSSSCASCMASSSSSSASASRSRSGGAGGASASRSASAGSVAS 737  
616 CCCCTCGCTCGGAGATCCCAAGTGGACCTGAGGTGCTTCACCTGCTGACGCA 675  
736 ASSSSSSSSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSV 677  
676 CAGTCAGAGAGTGCAAGCCCGGAGAGAGAGCCCGGCGGACCAAGCCAGGAGA 735  
676 SNAWSAARSSSSSSSSSSSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSS 617  
736 CCCCGCTCCCGCGGAGCTTAAAGTCCGCGGAGAGAGAGTAAAGTCCACCTGGAG 795  
616 SVASASGMSVSSSVSSGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 557

QY 796 CCTAGCCGACGACGACGAC 819  
DB 556 SSAAAAAASCVASCGMCGSKS 533

RESULT 51  
LOCUS BG511885  
DEFINITION BG511885 587 bp mRNA linear EST 28-MAR-2001  
dad29d08.y1 wellcome CRC PCS107 tropicalis St10-12 Xenopus  
tropicalis cDNA clone IMAGE:4440999 5' similar to TR:095854 095854  
TRANSCRIPTIIONAL INTERMEDIARY FACTOR 1 ALPHA. ;, mRNA sequence.

ACCESSION BG511885  
VERSION BG511885.1 GI:13482542  
KEYWORDS EST  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis

REFERENCE Clifton, S., Johnson, S. L., Blumberg, B., Song, J., Hillier, L.,  
Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y.,  
Peterson, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,  
Waterson, R. and Wilson, R.  
Xenopus tropicalis; Xenopus; Silurana.

TITLE Washu Xenopus EST project, 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Sandy Clifton, Ph.D.  
Washu Xenopus EST project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Library constructed by A. Zorn and J. Mason (Wellcome/CRC  
Institute). DNA Sequencing by: Washington University Genome  
Sequencing Center  
Clone distribution: Xenopus clones from this library are available  
through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 502.

FEATURES  
source  
1. 587  
Location/Qualifiers  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8354"  
/clone="IMAGE:4440999"  
/issue\_type="whole embryo, stages 10-12"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Wellcome CRC PCS107 tropicalis St10-12"  
/note="Vector: PCS107; Site 1: NotI; Site 2: EcoRI; cDNAs  
were oligo-dT primed and directionally cloned. Average  
insert size 1.5 kb, range 0.5-4 kb. Library constructed by  
A. Zorn and J. Mason (Wellcome/CRC Institute)."

## ORIGIN

Query Match 4.1%; Score 62.6; DB 4; Length 587;  
Best Local Similarity 61.2%; Pred. No. 0.0022;  
Matches 101; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 528 AATGAGACGAGTGTCCGTGTGCGGAGCGCGGAGCTCATCTGTGACGCGCTCG 587  
DB 423 AATGAGACGAGTGTGCTGTATGTATTAATGCGGGGAGCTGCTTTGCTGTGAAATGT 482  
QY 588 CCTCGGGCTTCCACCTGCGCTGTCCCTCCGCTCCGGAGATCCCACTGGGAC 647  
DB 483 CCCAAAGTTTCCACCTGTGTGTGCGACGTCCTACCTGATGAACCTTCCCACTGTGAA 542  
QY 648 TGGAGGTGCTCCAGCTGCTGCGGACGACGACGACGACGAGGTGAG 692  
DB 543 TGGATCTGACCTTTTGGCCGGGACCTGTCCAGACGACGAGGTGAG 587

RESULT 52  
LOCUS CN526117  
DEFINITION CN526117 677 bp mRNA linear EST 29-APR-2004  
UI-M-HNO-coc-j-05-0-UI-r1 NIH\_BMAP\_HNO Mus musculus cDNA clone  
IMAGE: 30654076 5', mRNA sequence.

ACCESSION CN526117  
VERSION CN526117.1 GI:46853770  
KEYWORDS EST  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouse1.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

FEATURES  
source  
1. 677  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30654076"  
/issue\_type="Upper Head"  
/dev\_stage="9.5-10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP HNO"  
/note="Organ: Head; Vector: pYX-Anc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Anc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CCAAGTGAAT. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 4.0%; Score 62.2; DB 7; Length 677;  
Best Local Similarity 66.7%; Pred. No. 0.0028;  
Matches 88; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 533 GAGACAGTGTCCGTGTGCGGAGCGCGGAGCTCATCTGTGACGCGCTCG 592  
DB 506 GGAATTTCTGTCCGCTGTGAAAGAGCGCGGAGCTCCTGTGTGACATATGCCCTTC 565  
QY 593 GGCCTTCCACCTGCGCTGTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTGAG 652  
DB 566 TTTCCTACCATCATCATGCTGTAACCCCGCTGCGACAGATCCCAAGGGAATGACT 625  
QY 653 GTGCTCCAGCTG 664  
DB 626 CTGTCCCGCTG 637

RESULT 53  
CN523529

LOCUS CB523529 680 bp mRNA linear EST 09-JUL-2003  
 DEFINITION UI-M-GHO-cer-1-07-0-UI.r1 NIH\_BMAP\_GHO Mus musculus cDNA clone  
 IMAGE:6843512 5', mRNA sequence.  
 ACCESSION CB523529  
 VERSION CB523529.1 GI:29356884  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 680)  
 NIH-MGC <http://imgc.ncl.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: PYX-5.  
 FEATURES  
 Location/Qualifiers  
 1..680  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6843512"  
 /issue\_type="Whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1ib="NIH BMAP GH0"  
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGACTGAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN  
 Query Match 4.0%; Score 62.2; DB 6; Length 680;  
 Best Local Similarity 66.7%; Pred. No. 0.0028; Indels 0; Gaps 0;  
 Matches 88; Conservative 0; Mismatches 44;  
 QY 533 GGACGAGTGTGCGGTGTGCGAGCGCGGAGCTCATCTGCTGTGACGCGCTGCGCTCG 592  
 DB 482 GGAATCTCTCGCTGTGCAAGAGCGCGGAGCTCTGTGCTGTGACACATGCCCTTC 541  
 QY 593 GGCCTTCACCTGCGCTGTGCTGTGCTGTGCTGTGCGGAGATCCCAAGTGGAGCTGGAG 652  
 DB 542 TTCTTACCAATTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGGAGATGGCT 601  
 QY 653 GTGCTCCAGCTG 664  
 DB 602 CTGTCCCGCTG 613

RESULT 54  
 BU056500/c 751 bp mRNA linear EST 26-AUG-2002  
 LOCUS BU056500

DEFINITION UI-M-F00-cab-h-16-0-UI.r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
 IMAGE:6409023 5', mRNA sequence.  
 ACCESSION BU056500  
 VERSION BU056500.1 GI:22496577  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 751)  
 NIH-MGC <http://imgc.ncl.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
 Seq primer: PYX-5.  
 FEATURES  
 Location/Qualifiers  
 1..751  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6409023"  
 /issue\_type="Whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1ib="NIH BMAP F00"  
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN  
 Query Match 4.0%; Score 62.2; DB 5; Length 751;  
 Best Local Similarity 66.7%; Pred. No. 0.0028; Indels 0; Gaps 0;  
 Matches 88; Conservative 0; Mismatches 44;  
 QY 533 GGACGAGTGTGCGGTGTGCGAGCGCGGAGCTCATCTGCTGTGACGCGCTGCGCTCG 592  
 DB 630 GGAATCTCTCGCTGTGCAAGAGCGCGGAGCTCTGTGCTGTGACACATGCCCTTC 571  
 QY 593 GGCCTTCACCTGCGCTGTGCTGTGCTGTGCTGTGCGGAGATCCCAAGTGGAGCTGGAG 652  
 DB 570 TTCTTACCAATTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGGAGATGGCT 511  
 QY 653 GTGCTCCAGCTG 664  
 DB 510 CTGTCCCGCTG 499

RESULT 55  
 AG561783 776 bp DNA linear GSS 05-JUN-2004  
 LOCUS AG561783

```

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-481P17.TU, genomic survey
sequence.
ACCESSION AG561783
VERSION AG561783.1 GI:48322481
KEYWORDS Mus musculus molossinus
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE BAC end Sequences of Library MSMg01
JOURNAL Unpublished
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
JOURNAL Direct Submmission
TITLE 2 (bases 1 to 776)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kunya Abe (abe@rtc.riken.jp). Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TU
LIBRARY Vector : pBAC3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source Location/Qualifiers
1.776
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-481P17.TU"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC library"
ORIGIN
Query Match 4.0%; Score 61.8; DB 9; Length 776;
Best Local Similarity 61.8%; Pred. No. 0.0035;
Matches 118; Conservative 0; Mismatches 67; Indels 6; Gaps 1;
QY 731 GGAGACCCCGCTCCCGGCTTAACTGCGGGAGAGAGAGTAAAGTCCACTCG 790
DB 62 GGAATTCCAGATCTCTGGAGCTGAGCTTCAAGAAAACAGAGGCCCGTCCAG 121
QY 791 GGAACCCCTAGACCGGATGACACGACTTGTCTCAAGACACTGCGGCTCCGCTTC 850
DB 122 GAGGCCCAAGCAGCTGATGCTGTCACATATGGAACCTGCGCCCAACACC 181
QY 851 TGCAGCCCCGCTGCGAGGTGAGTCTCTCGGCCCTGCACCCCCCTACTGTGTGGGCTC 910
DB 182 TGCAGCTCTCTG-----CTGAGGCTTCAGCAGCTGTGCTCTGAGTCTGGGAGC 235
QY 911 TGAAGGTCAAGC 921
DB 236 TGAAGGAGCAGC 246
RESULT 56
LOCUS CE173623 262 bp DNA linear GSS 25-SEP-2003
DEFINITION tigr-gss-dog-17000326704476 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION CE173623

```

```

VERSION CE173623.1 GI:35316113
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1
AUTHORS Kirnesh,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Fop,M., Wang,W., Frazer,C.M. and
Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirnesh EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirnesh@tigr.org
Class: Shotgun.
FEATURES
source Location/Qualifiers
1.262
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BseXI; Libraries were prepared from peripheral blood"
ORIGIN
Query Match 4.0%; Score 61.6; DB 9; Length 262;
Best Local Similarity 66.7%; Pred. No. 0.004;
Matches 88; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 533 GGACGAGTGGCCGGTGTGTCGGAGCGGGAGGCTCATCTGTCGACGGCTCGC 592
DB 110 GAGGTTCTCCCGCGTGTGCAAGACGCGGAGAGCTCTGCTGCGACGCTGCCCTTC 169
QY 593 GGCCTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGCGAGATCCCACTGAGAG 652
DB 170 CTCCTACACACCTGCACTGCTTCACCCCGCGTGCAGATCCCAAGGATGCT 229
QY 653 GTGCTCAGCTG 664
DB 230 CTGCCCGCGCTG 241
RESULT 57
LOCUS BE234617 434 bp mRNA linear EST 10-JUL-2000
DEFINITION 141780 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE234617
VERSION BE234617.1 GI:9019335
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
REFERENCE 1 (bases 1 to 434)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug,S.C., Smith,T.P.L., Frøking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R.,
Quackenbush,J. and Keefe,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE 22213769
PUBMED 12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

```











Email: asimpor@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4-BN0090-020  
 600-236-g06&tc3=2000-06-02&ct=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 359.

## FEATURES

## source

Location/Qualifiers

1..359  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="adult"  
 /clone\_lib="BN0090"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 3.9%; Score 60; DB 2; Length 359;  
 Best Local Similarity 61.5%; Pred. No. 0.0082;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGAGCTCATCTGTCAGCGCTGCCCTC 591  
 |||||  
 DB 187 AGGACTATTGCGAGGTGTGCGACGAGCGGTGATCATCTGTGTGATACCTGTCC 128  
 |||||  
 QY 592 GGGCCTTCCACCTGCGCTGCTGCTCCCTCCGTCGCGAGATCCCAAGTGGACCTGGA 651  
 |||||  
 DB 127 GTGCTTACCAACAGTGTGCTGCTGATCCGACATGAGAGGCTCCCGAGGGCAAGTGA 68  
 |||||  
 QY 652 GGTGCTCCAGCTGCTGCTGCGGCAACAGTCCAGAGG 687  
 |||||  
 DB 67 GCTGCCACACTGCGAGAGGAGGATCCAGTGGG 32  
 |||||

RESULT 66  
 CF137810 461 bp mRNA linear EST 09-SEP-2003  
 LOCUS  
 DEINITION  
 IMAGE:3093236 5', mRNA sequence.  
 CF137810.1 GI:3253254  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 461)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newcom Road, 4156 MEBR, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Louis Staudt  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/humanfl.html  
 Seq primer: pyx-5.

## FEATURES

## source

Location/Qualifiers

1..461  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:3093236"  
 /issue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LTI)"  
 /clone\_lib="NIH MGC 50"  
 /note="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

## ORIGIN

Query Match 3.9%; Score 60; DB 6; Length 461;  
 Best Local Similarity 61.5%; Pred. No. 0.0084;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGAGCTCATCTGTCAGCGCTGCCCTC 591  
 |||||  
 DB 122 AGGACTATTGCGAGGTGTGCGACGAGCGGTGATCATCTGTGTGATACCTGTCC 181  
 |||||  
 QY 592 GGGCCTTCCACCTGCGCTGCTGCTCCCTCCGTCGCGAGATCCCAAGTGGACCTGGA 651  
 |||||  
 DB 182 GTGCTTACCAACAGTGTGCTGCTGATCCGACATGAGAGGCTCCCGAGGGCAAGTGA 241  
 |||||  
 QY 652 GGTGCTCCAGCTGCTGCTGCGGCAACAGTCCAGAGG 687  
 |||||  
 DB 242 GCTGCCACACTGCGAGAGGAGGATCCAGTGGG 277  
 |||||

RESULT 67  
 BU429642 544 bp mRNA linear EST 09-SEP-2002  
 LOCUS  
 DEFINITION  
 IMAGE:3064481 5', mRNA sequence.  
 BU429642.1 GI:22768129  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 544)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Eco RI site shown at the beginning of the sequence.  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: M.B. Soares Lab  
 CDNA Library Arrayed by: M.B. Soares Lab  
 DNA Sequencing by: M.B. Soares Lab  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html  
 Seq primer: M13 Forward  
 Location/Qualifiers

## FEATURES

## source

1..544  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:3064481"  
 /issue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LTI)"  
 /clone\_lib="NIH\_MGC\_50"



Oy	652	GGTGTCCAGTCGCTGCAGCACAATGCCAGAG	687
Dd	203	GCTGCCCACTACCGCAGAAGAGCATCCACTGGG	238
RESULT 70	Bg329665	699 bp	mRNA linear EST 27-FEB-2001
LOCUS	602422726P1 NIH_MGC_16 Homo sapiens	CDNA clone IMAGE:4558915 5'	
DEFINITION	MGC sequence.		
ACCESSION	Bg329665		
VERSION	Bg329665.1 GI:13136103		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 699) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@bs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU) DNA sequencing by: Incyte Genomics, Inc. Gene distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNU at: image.lnl.gov Plate: LNCMI265 row: h column: 20 High quality sequence stop: 650.		

```

FEATURES
source
    Location/Qualifiers
        1..699
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4558915"
            /tissue_type="retinoblastoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_16"
            /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Library constructed by Ling Hong
            in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
            Note: this is a NIH_MGC Library."

```

Query Match	3.94%	Score 60;	DB 4;	Length 699;
Best Local Similarity	61.54%	Pred. No. 0.0088;		
Matches	96;	Conservative	0;	Mismatches 60; Indels 0; Gaps 0;
OY	532	AGACGAGTGTGCGGTGTCTGCGGACGGCGGAGACTCATCTGTGACGCGTGCCTC	591	
Db	35	AGGACTATTGCGAGGTGTGCGCAGCAGCGGTGAGATCATCTGTGTATACCTGTCCC	94	
OY	592	GGGACCTTCACCTGGCCCTGCTGCCCTCCCTCCGGCTCCGGAGAGTCCCGCAGTGGGACCTGGA	651	
Db	95	GTGCTTACCAATGTCGTCTGCTCGATCCCGACATGAGAGAGAGGCTCCGAGGGCAAGTGA	154	
OY	652	GGTGTCTCAAGTGTGCTCTGACGGCAACAGTCCACGAGAG	687	
Db	155	GCTGCCCCACACTGCGAGAGAGAGGCAATCCAGTGG	190	
RESULT 71				
BE258675				
LOCUS	BE258675	700 bp	mRNA	linear
DEFINITION	601115344F1 NIH_MGC_16 Homo sapiens CDNA clone IMAGE:3356962 5',			EST 13-JUL-2000
ACCESSION	BE258675			

VERSION	BE258675.1	GI:9129168
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 700)	
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph. D. Email: <a href="mailto:cgabds-remail.nih.gov">cgabds-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">image.lnl.gov</a> plate: LNCM163 row: g column: 11.	
FEATURES		
Source	1..700	
	Location/Qualifiers	

```

FEATURES
    source
        location/Qualifiers
            1..700
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3356962"
                /tissue_type="retinoblastoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_1lb="NH_MGC_16"
                /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GCGACGCG(G). Library constructed by Ling Hong
                in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
                Note: This is a NIH MGC Library."

```

	Query Match	Similarity	61.9%;	Score	60;	DB 2;	Length	700;		
	Best Local	Similarity	51.5%;	Pred.	No.	0.0088;				
	Matches	96;	Conservative	0;	Mismatches	60;	Indels	0;	Gaps	0;
OY		532	AGACGAGTGTCCGTGTGTGGGACGCCGGAGCTCATCTGCTGTGACCGCTGCCTC	591						
Db		70	AGACTATTTCGAGAGTGTCGCAGCAAGCCGATGAGATCATCTGTGTGATTA	129						
OY		592	GGGCGCTTCCACCTGAGCGCTGCTGCCCCCTCGCGCTCCGGGAGATCCCAGTAGGGCACTGGA	651						
Db		130	GTGCTTTACCATGATGATGTCTGCTGTGATTCGCACATGAGAAGAGCTCCGAGGGCAATGGA	189						
OY		652	GGTGCTTCAGCTGCTGTCGACGACAACACTGTCAGGAGC	687						
Db		190	GCTGCCACACACTGCGAAGAAAGGACATCCAGTAGGG	225						
RESULT 72										
LOCUS										
BE176589										
DEFINITION		RC3-HT0585-010400-023-c04 HT0585 Homo sapiens cDNA,	731 bp	mRNA	linear	EST 21-JUN-2000				
ACCESSION		BE176589								
VERSION		BE176589.1	GI:8639318							
KEYWORDS		EST.								
SOURCE		Homo sapiens	(human)							
ORGANISM		Homo sapiens								
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
		1 (bases 1 to 731)								
REFERENCE		Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Britones,M.R.,								
AUTHORS		Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin S., Costa,F.F.,								
		Goldman,G.H., Carvalho,A.P., Matsukuma,A., Bata,G.S., Simpson,D.H.,								
		Brunstein,A., deOliveira,P.S., Bucher,F., Jongeneel,C.V.,								
		O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.U. and								
		Simpson,A.J.								



TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
PUBMED	10737800
COMMENT	Contact: Simpson A.J.G.

## ORIGIN

Query Match	3.9%;	Score 60;	DB 2;	Length 731;
-------------	-------	-----------	-------	-------------

OY 532 AGGACGAGTGTGCGCGTGTGTCGGGAGCGGGGAGCTCATCTGCGTGAACGGCTGACCTC 591  
 Db 376 AGGACATATTGCGAGGTGTGTCGACGACAAAGGCGGTGAATCATCTGTGTATATACCTGTCCC 435  
 OY 592 GGGCGCTTCCACCTCGGGCTGCTGCTCTCCCTCCGCTCCGGAATCCCAAGTGGACCTGGA 651  
 Db 436 GTGTGTTTACCACTAGTGTCTGTGCTGATCTCCGACATGAGAGAGGCTCCCGAGGCAAGTGA 495  
 OY 652 GGTGCTTCCAGCTGCTCTGCAAGGCACAATCCAGAGAG 687  
 Db 496 GCTGCCCCACCTGCGAAGAGAAAGGATCACTCAATGGG 531

RESULT 73	BE297517	750 bp	mRNA	linear	EST 20-JUL-2000
LOCUS	BE297517				
DEFINITION	601118063p1 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:3533391.5,				
ACCESSION	BE297517				
VERSION	BE297517.1	GI:9181002			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 750)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	unpublished (1999)				
COMMENT	Contact: Robert Strusberg, Ph.D.				
	Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a>				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Ling Hong/Rubin Laboratory				

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LNCM10 row: n column: 16  
High quality sequence stop: 550.  
location/Qualifiers  
1. 750

## ORIGIN

Query Match	3.9%;	Score 60;	DB 2;	Length 750;
Best Local Similarity	51.5%;	Pred NC	0.0089.	

QY	532	AGAGCAAGATGTGCGCGTGTGTGAGGACGGCGGGGAGCTCATCTGTGCTGTACAGGCTGCCTC	591
Db	6	AGCACTATTGGAGTGGTGTGCGCAGCAAGCGGTGTGATCATCTGTGTATTACTGTCCC	65
QY	592	GGGCGCTTCCACTGTGGCGCTGTCCCTCCGCTCCGGGAGATCCCAAGTGGACCTGGA	651
Db	66	GTCCTTACCAATGTCGTCTGTGATCCCAATGTGAGAAAGGCTTCCAGGCAAGTGA	125
QY	652	GGTGTCTCAGTGTGCTGTGAGGCAACAGTTCAGAGG	687
Db	126	GCTGCCCACTGCAGAAAGGACATTCAGTGG	161

FEATURES	source
ACCESSION	BE259148
LOCUS	BE259148
DEFINITION	601108075F01 NIH_MGC_16 Homo sapiens CDNA clone IMAGE:3344251 5', mRNA sequence.
VERSION	BE259148
KEYWORDS	BE259148.1 GI:9129658
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 817)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Straubeberg, Ph.D. Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a> Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMU at: <a href="http://image.llnl.gov">image.llnl.gov</a> Plate: L10CM130 row: e column: 20. Location/Qualifiers 1..817
	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3344251"



QY 652 GGGGCTCCAGCTGCTGCGAGGACGACAGTCCAGGAGG 687  
| | | | | | | | | | | | | | | | | | | | | |  
Db 124 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 159  
| | | | | | | | | | | | | | | | | | | | | |

RESULT 77  
LOCUS BG394104  
DEFINITION 602456272F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:4579048 5',  
RNA sequence.  
ACCESSION BG394104  
VERSION BG394104.1 GI:13287552  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: L16M1235 row: 0 column: 17  
High quality sequence stop: 635.  
Location/Qualifiers  
1. 974  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4579048"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH\_MGC\_16"  
/note="Organ: eye; Vector: pOT7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."

ORIGIN

Query Match 3.9%; Score 60; DB 4; Length 974;  
Best Local Similarity 61.5%; Pred. No. 0.0091;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCTGCTGCGAGCGGGAGCTCATCTGCTGACGAGTGCCTTC 591  
| | | | | | | | | | | | | | | | | | | | | |  
Db 35 AGGACATATTGCGAGGTGTGCGACGAGGGGTGATCATCTGTGTACTGTCCTCC 94  
| | | | | | | | | | | | | | | | | | | | | |

QY 592 GGGCCTTCACCTGCGCTGCTGCTCCCTTCGCTCCGGAGATCCCAAGTGGACCTGGA 651  
| | | | | | | | | | | | | | | | | | | | | |  
Db 95 GTGCTTACCACTGCTGCTGCTGATCCGACATGGAAGGCTCCCGAGGGCAAGTGA 154  
| | | | | | | | | | | | | | | | | | | | | |

QY 652 GGGGCTCCAGCTGCTGCGAGGACGACAGTCCAGGAGG 687  
| | | | | | | | | | | | | | | | | | | | | |  
Db 155 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 190  
| | | | | | | | | | | | | | | | | | | | | |

RESULT 78  
LOCUS BM457082  
DEFINITION 989 bp mRNA linear EST 05-FEB-2002  
AGENCOURT\_6406806 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:5583326  
5', mRNA sequence.  
ACCESSION BM457082

VERSION BM457082.1 GI:18506122  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L16M12346 row: d column: 15  
High quality sequence stop: 658.  
Location/Qualifiers  
1. 989  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5583326"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

ORIGIN

Query Match 3.9%; Score 60; DB 4; Length 989;  
Best Local Similarity 61.5%; Pred. No. 0.0091;  
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCTGCTGCGAGCGGGAGACTCATCTGCTGACGAGTGCCTTC 591  
| | | | | | | | | | | | | | | | | | | | | |  
Db 38 AGGACATATTGCGAGGTGTGCGACGAGGGGTGATCATCTGTGTACTGTCCTCC 97  
| | | | | | | | | | | | | | | | | | | | | |

QY 592 GGGCCTTCACCTGCGCTGCTGCTCCCTTCGCTCCGGAGATCCCAAGTGGACCTGGA 651  
| | | | | | | | | | | | | | | | | | | | | |  
Db 98 GTGCTTACCACTGCTGCTGCTGATCCGACATGGAAGGCTCCCGAGGGCAAGTGA 157  
| | | | | | | | | | | | | | | | | | | | | |

QY 652 GGGGCTCCAGCTGCTGCGAGGACGACAGTCCAGGAGG 687  
| | | | | | | | | | | | | | | | | | | | | |  
Db 158 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 193  
| | | | | | | | | | | | | | | | | | | | | |

RESULT 79  
LOCUS BF309509  
DEFINITION 1011 bp mRNA linear EST 21-NOV-2000  
601892016F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4137843 5',  
RNA sequence.  
ACCESSION BF309509  
VERSION BF309509.1 GI:11256902  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incycle Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLM1045 row: P column: 04  
 High quality sequence stop: 697.  
 Location/Qualifiers

## FEATURES

source

1..1011  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4137843"  
 /issue\_type="rhodomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /lab\_lib="NIH MGC 17"  
 /note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;  
 Site 2: XhoI; CDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAAGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 3.9%; Score 60; DB 2; Length 1011;  
 Best Local Similarity 61.5%; Pred. No. 0.0092;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTCGGGAGCGGAGGAGCTCATCTGTCGACGCTGCCCTC 591  
 |||||  
 DB 26 AGACTATTGCGAGGTGTCGACGACGAGCGGTGATCATCTGTGTATCTGTCC 85  
 592 GGGCTTCACACTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCACTGAGCACTGGA 651  
 |||||  
 DB 86 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAAAGCTCCGAGGCGAATGGA 145  
 |||||  
 QY 652 GGTGCTCCAGCTGCTGCGGACGCAACAGTCCAGAGG 687  
 |||||  
 DB 146 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 181  
 |||||

RESULT 80  
 BQ213031 1365 bp mRNA linear EST 02-MAY-2002  
 LOCUS BQ213031  
 DEFINITION AGENCOURT 7591269 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6067251  
 5', mRNA Sequence.  
 ACCESSION BQ213031  
 VERSION BQ213031.1 GI:20393828  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1365)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM13346 row: h column: 04  
 High quality sequence stop: 446.  
 Location/Qualifiers

## FEATURES

source

1..1365  
 /organism="Homo sapiens"

/mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6067251"  
 /issue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /lab\_lib="NIH MGC 93"  
 /note="Organ: testis; Vector: pCMV-Sport6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH MGC library."

## ORIGIN

Query Match 3.9%; Score 60; DB 5; Length 1365;  
 Best Local Similarity 61.5%; Pred. No. 0.0095;  
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTCGGGAGCGGAGGAGCTCATCTGTCGACGCTGCCCTC 591  
 |||||  
 DB 64 AGACTATTGCGAGGTGTCGACGACGAGCGGTGATCATCTGTGTATCTGTCC 123  
 592 GGGCTTCACACTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCACTGAGCACTGGA 651  
 |||||  
 DB 124 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAAAGCTCCGAGGCGAATGGA 183  
 |||||  
 QY 652 GGTGCTCCAGCTGCTGCGGACGCAACAGTCCAGAGG 687  
 |||||  
 DB 184 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 219  
 |||||

RESULT 81  
 BF847408 426 bp mRNA linear EST 16-JAN-2001  
 LOCUS BF847408  
 DEFINITION PM4-EN0063-301000-001-e05 EN0063 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF847408  
 VERSION BF847408.1 GI:12234558  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 426)  
 Dias Neto, F., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Coêta, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 JOURNAL MEDLINE  
 PUBLISHED 20202663  
 10737800  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the PABSP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
<http://www.ludwig.org.br/serlpts/gethtml2.pl?l=PM4&t2=PM4-EN0063-301000-001-e05&t3=2000-10-30&t4=1>  
 Seq primer: puc 18 forward  
 High quality sequence stop: 14  
 Location/Qualifiers

## FEATURES

source

1..426  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

```

/dev stage="Adult"
/clone.lib="EN0063"
/note="Torgan: Lung normal, Vector: puc18, Site_1: Smal,
Site_2: Sma; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

Query Match	3.9%	Score 59.8;	DB 2;	Length 426;
Best Local Similarity	58.9%;	Pred. No. 0.0093;		
Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0;

Qy	AGACAGAGTGTGACCGTGTGTGGGACCGGCGGGAGCTCATCTGGCTGTGACGAGCTGCCTC	591
Db	AGCATTACTGTGAGAGTGTGCCACGAGGTGGGAAATATTCTGTGTGTACACTGCTCCTC	306
Qy	GGGCTTTCANCTGGCTGCTGTTCCTTCGCTCCGGGAAATATCCCACTGTGGACCTGTGA	651
Db	GTGCTTACACCTGTGTCTGTCTTGAATCTTGAGCTTGAACGGGCTTCAGAGGGCAAAATGTGA	366
Qy	GGTCTCTCAGTGTGCTGTGAGGCAACAGTCCAGAGAGTGCACGCCCGGCGACAGGA	706
Db	GCTCCCTCTCATGTGTGAGAAAGGAGGGGTCTCAGTGTGGAGGCCCAAGAGGAGGAAACA	421

RESULT	82				
LOCUS	BU193516				
DEFINITION	BU193516	885 bp	mRNA	linear	EST 04-SEP-2002
ACCESSION	AGNCOCURT_7840454	NIH_MGC_67	Homo sapiens	CDNA clone IMAGE:6146430	
VERSION	BU193516				
KEYWORDS	BU193516.1	GI:22707500			
SOURCE	EST.				
ORGANISM	Homo sapiens				
	Homo sapiens (human)				

REFERENCE 1 (pages 1 to 885)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
 Plate: LLM13474 row: k column: 07  
 High quality sequence stop: 724.

FEATURES	Location/Qualifiers
source	1. .885

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6146430"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NH_MSC_67"
/notes="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 1.75 kb. Library constructed by Life
Technologies."

```

Query Match	3.9%	Score 59.8;	DB 5;	Length 885;
Best Local Similarity	58.9%	Pred. No. 0.01;		
Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0;

QY	532	AGAGAGATGTGTCCGTGTGTGTGGGAGCGGCGGGAGCTACTCTGCTGTATACGGTGTGCCCTTC	551
Db	383	AGGATTACTGTGAGGTGTGTGCCAGCAGGGGTGGGAAATTATTTCTGTGTGCACTCTGCCCTTC	442
QY	592	GGGCTTTTCACCTGTGGCTTGTGCTTGTCCCTCCGCTCCGGAGATATCCCAATGGGACCTTGA	651
Db	443	GTGCTTACACACTTCGTCTGTCTTGTATTCCTGACCTTGAACCGGGCTCCAGAGGGCAAAATGGA	502
QY	652	GGTGTCTCAGGCTGCTGCAGGCAACATTCACAGAGGTGCAGGCCCCGGCAGCAGGA	706
Db	503	GCTGCTCTTCACCTGTGAGAGGAGGGGGGTTCATGTGGAGGCCCAAGAGGAAAGAAAGA	557

RESULT 83					
BUS27069					
LOCUS	BUS27069	918 bp	mRNA	linear	EST 13-SEP-200
DEFINITION	AGENCOURT_10155962 NIH_MGC_101 Homo sapiens CDNA clone				
	IMAGE:6536820 5', mRNA sequence.				

SOURCE ORGANISM	REFERENCE
Homo sapiens (human)	1 (bases 1 to 918)
Homo sapiens	
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 918)	NIH-MGC	<a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		
	Contact: Robert Strausberg, Ph. D.			
	Email: <a href="mailto:cgabbs-r@mail.nih.gov">cgabbs-r@mail.nih.gov</a>			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Rubin laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LNL at:			
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>			
	Plate: LTCM2698	row: m	column: 12	
	High quality sequence	stop: 666.		

FEATURES	Location/Qualifiers
source	1. .918

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6536820"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC 101"
/notes="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

```

ORIGIN

Query Match	3.9%	Score 59.8	DB 5	Length 918
Best Local Similarity	58.9%	Pred. No. 0.01		
Matches 103; Conservative	0	Mismatches 72	Indels 0	Gaps 0

OY  
Db

532 AGGACGAGTGTCCCGTGTCGGAGCAGCGGGGAGCTCATCTGCCTGTGAACGGGTGCCCTC 591  
|||||  
145 AGGATTACTGTAGGCTGTCCAGCAGCGGTGGGAAATTATCTGTGTGAACACTGCCCTC 204  
|||||

```

Oy      592 GGGCTTCCACCTGCGCTGCGCTGCTGCCGCTCCGGAGATCCCAAGTGGGACCTCGGA 651
      |||||
Db      205 GTGCTTACCACCTCGTCTGCTTGTATCTTGAAGCTTGAACCGGGCTCCAGAGGGCAATGGA 264
      |||||
Oy      652 GGTGCTCCAGCTGCTCGACAGGCACAGTCCAGAGAGGTGACGCCCGGGACGAGAGA 706

```

Db	265	GCTGCCCTC	ACTGTGTA	GAAAGAG	GGGGGTCT	CACTGGAG	CCCAAGAA	GAGAAGA	319
RESULT 84									
BUS28543									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
COMMENT									
FEATURES									
source									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
3.9%									
Score 59.8									
DB 5									
Length 943									
58.9%									
Pred. No. 0.01									
0									
Mismatches 72									
Indels 0									
Gaps 0									
532									
AGAGCAGATGTC									
CGCTGTGTCTG									
CGGAGCGCGG									
AGAGCTCATCT									
GTGTGAGCGG									
CTGCCTC									
591									
178									
AGGATTACTGT									
GAGGTGTGCC									
AGAGGTGGGA									
ATTATCTGTG									
TACACCTG									
CCCTC									
237									
592									
GGGCTTTCAC									
CTGTGCGCT									
CTGTCTCCCT									
CTCCGGAGAT									
CTCCCACTG									
GGAGCTTGA									
651									
238									
GTGCTTACCA									
CTCTCTGTG									
CTTGTAGCTT									
GTACCCGGCT									
TCAGAGGCA									
AAATGGA									
297									
652									
GATGCTCAG									
AGCTGTGCA									

[illegible]

QY	1072	CAGAGACGTAACCCGAGCCCCCTGTGAGGGGGGTGTGTGGCCCCCAGCCGCCGCCGCTGG	1131
Db	554	CCGGGGGCCCCCGCCGCGCGCGCGCGCGCGCGCCCGCGGGCCCCCGCCCGGGG	495
QY	1132	CCCTGTGGGCTTGCAGAGATGACATGTGCAGTCAAGCCCCGCTTGTGCACAGGATGACC	1191
Db	494	CGCCCGGCGGGCCCCCCCCCCCCCCCCCCCCCGGCCCCCCCCCGGCNCCCCC	435
QY	1192	TGGAGTCCCTTCTGAGCGACACCTTGCATGGGATCTGTGCAGTGGGCCATCCAGACA	1251
Db	434	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNCCCCCNCGCCCCCCCCNCC	375
QY	1252	TGGCCCGTCCGGCGGCCCCCTTCCCTCTGACCCAGATGACCGGGACATGCACTCTG	1311
Db	374	CCCCCCCCCCCCCCCCCGCCNCCGCCNCCCCCCCCCNCCNCCCCCCCCCCCCCCCN	315
QY	1312	ATTGAGAGTGTCTGAGAGGACACCTCTCTTCTCATGCTCTGGAAAC	1357
Db	314	NNNNNNNNNNNNNNNNNNNNNNCCCACTCATGCTGTGTCAAGC	269

RESULT	86
B0672163	
LOCUS	B0672163 1132 bp mRNA linear EST 15-JUL-2002
DEFINITION	AGENCOURT_8302070 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274983 5', mRNA sequence.

ACCESSION	506/2163
VERSION	B0672163.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1132)  
NIH-MGC <http://mgc.ncbi.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [CGAPbs-remail.nih.gov](mailto:CGAPbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINTL at:  
<http://image.lnlnl.gov>  
Plate: LUCM2457 row: g column: 16  
High quality sequence stop: 306.

## FEATURES

### source

```

1. .1132
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6274993"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH MGC 102"
/notes="Organ: salivary gland; Vector: pOT7; Site_1: XhoI; Site_2: EcoRI. cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

```

## ORIGIN

Query Match	3.8%;	Score 58.8;	DB 5;	Length 1132;
Best Local Similarity	31.8%;	Pred. No. 0.017;		
Matches 276; Conservative	0;	Mismatches 592;	Indels 0;	Gaps 0;

38 GCGTTTGATCCAGGCTGTAACCGCTGCTCTCAGCTGGAGCCCGTGGGTGGGCCCGGCGCCC 97

Db	224	GCCTGGGCGCAGCGCAGACCGCGCTTCTCTCGCGGCGCGCGGCGGCGCTCGGAGC	283
Oy	98	CTGCTATAGCCAGAGGTCAGAGATCCATGGGATGCTCTCATCTTTCGTCCAG	157
Db	284	CGCCCGCGGGGTGGGTNN	343
Oy	158	CATGTTTCTTAATGGGGTAGAAGCAGTCCGAGAGACCTCCCTGGGCTGGCCCACT	217
Db	344	NN	403
Oy	218	GCCTGTGAGAAAGGTTCAITGGTTGGTACATTCCGGGCGCCCTGGAACGACGA	277
Db	404	NN	463
Oy	278	GCTTGCAGAAACC GG GTTTTCTTCCCAATAGGATGGCCCCGGGGGGTGTCTTTAG	337
Db	464	NN	523
Oy	338	ACCAGATGSAITGGGGAACAGTGTCTCAGGCGACAATTTCAGGCTCTGGCAGCATGGAGC	397
Db	524	GG	583
Oy	398	AGGCGAAGACTGGGGAGTTCAAGTACCAGAGATGCTCTGGGGGAGCTGTTTGGAA	457
Db	584	NNGG	643
Oy	458	GGAGTGGCTCTCAGAGGGGTGTCGACCCAGCCAGTCTGACATGGGCGTCTTGTCT	517
Db	644	GGGGGGGGGNGNNNNGG	703
Oy	518	GTGCGAGAGATAGGACGAGTGTGCTGTGTGCGGACCGCCGGAGACTCATCTTCTG	577
Db	704	NNGG	763
Oy	578	TGAGGCTGCTTGGGCTTTTCCACTTGGCTGCTGCTTCCCTCCGTCCTGGAGATCCC	637
Db	764	CCGG	823
Oy	638	CAGTGGGACTGAGAGTGTCTCAGCTGCTTGCAGGACAAAGTCCAGAGAGTGCAGCCCG	697
Db	824	GGCGCGGGGNNGCCCGCGGGGGGTCCCGCGGGCGCGCGGGCGCGCGCGCGCGCC	883
Oy	698	GGCAGAGAGCCCCGGCCCCAGAGGCCACCCGTGAGACCCCGTCCCCCGGGCTTAA	757
Db	884	GGGCC	943
Oy	758	GTCCGCGGGAGAGAGGTAAAGGTCCACTGGGGAAACCTCTACCGGCAATGACAGAC	817
Db	944	GGCCCCCGCGACCGCC	1003
Oy	818	TCTTGTCTACAGAGACTGTCGAGCTCGGCTTCTGACAGCCCCGCTCCAGGCTTGAATC	877
Db	1004	CGCNCGCGCCCCCGCC	1063
Oy	878	CTCGGCGCTGACCCCTCACTGTGTGTG	905
Db	1064	CGGGCGGGCGCNCCTGGTCCGCGCGGG	1091

RESULT	87
BM450237	
LOCUS	
DEFINITION	BM450237 1350 bp mRNA linear EST-05-FEB-2002 AGNCOURT 6393396 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528315 5' mRNA sequence.

REVISION	21110237
VERSION	BM450237.1
	GI:18499277

## KEYWORDS

**SOURCE**  
**ORGANISM**

REFERENCE 1 (bases 1 to 1350)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.



**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Straube, Ph.D.  
 Email: c9apbs-x@mail.nih.gov  
 Tissue Procurement: ATCC/CDT/DRP  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LLM12204 row: s column: 12  
 High quality sequence stop: 370.

**FEATURES**  
 source  
 Location/Qualifiers  
 1. 1350  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5528315"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1ib="NH MGC 72"  
 /note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."

**ORIGIN**

**Query Match** 3.8%; Score 58.6; DB 4; Length 1350;  
 Best Local Similarity 41.0%; Pred. No. 0.02;  
 Matches 267; Conservative 0; Mismatches 378; Indels 7; Gaps 1;

477 GTGCTGACCCCAAGCTGATGAGGCTCTTCTGTCAGAGAAATGAGAC 536  
 705 GGGGCCCCCGGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 764  
 537 GAGTGTGCGGTGTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 596  
 765 CGGCG 824  
 597 TTCACCTGAGCTGCTGCTCCCTGCTCCGCTCCGCGGAGATCCCAAGTGAAGTGC 656  
 825 CG 884  
 657 TCCAGCTGCTGAGGCAAGTTCAGAGAGTGCAGCCCGGCGAGAGAGCCCGGCG 716  
 885 CCG 937  
 717 CAGAGCCCAAGCTGAGAGCGCGCTCCCGCGGCGCTTAAAGTCCGCGGAGAGAGTA 776  
 938 CCG 997  
 777 AGAGTTCACCTGAGGAGACCCCTAGCGGCGATGAGACAGCTTTGTACAGACACTG 836  
 998 GGGGGGCG 1057  
 837 CGGCGCTCGCGCTTCTGACAGCCCGCTGCAAGTGTGAATCTCTGCGCCCTGACCCCTTA 896  
 1058 NNCCNCCCCCCCCCNNGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1117  
 897 CTGTGTGTGGGTCTGAGGATGAGAGAACTGTGCTCTGTGTGTGTGTGTGTGTGTG 956  
 1118 NCCNCCG 1177  
 957 GGAGATGTATCGGACGTGCTGCGGTGTACTCACTGCGCGCTGCTTCACTGAGCGCTGC 1016  
 1178 CCGGNNNGGNNCCG 1237  
 1017 CACTTCCAGCGCGGACCTTCGCGCGCGCGGAGCGGCGCTGCGCTGAGATCTGCTCAGA 1076  
 1238 CCG 1297  
 1077 GACGTGACCCCAAGCCCTGTGAGAGGAGGTGTGAGCGCGCGCGCGCGCGCGCG 1128

Db 1298 NCGCGGCG 1349

**RESULT 88**  
**LOCUS** CB606244 416 bp mRNA linear EST 16-MAY-2003  
**DEFINITION** AMGNNUC:SRPB2-00055-Cl1-A srpb2 (10220) Rattus norvegicus cDNA  
 clone srpb2-00055-cl1 5', mRNA sequence.  
**ACCESSION** CB606244  
**VERSION** CB606244.1 GI:29545856  
**KEYWORDS** EST.  
**SOURCE** Rattus norvegicus (Norway rat)  
**ORGANISM** Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

**REFERENCE**  
 1 (bases 1 to 416)  
 Amgen EST Program.  
**AUTHORS** Amgen Rat EST Program  
**TITLE** Unpublished (2003)  
**JOURNAL** Contact: Dan Fitzpatrick  
**COMMENT** Amgen, Inc  
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00055 row: c column: 11.  
 Location/Qualifiers  
 1. 416

**FEATURES**  
 source  
 Location/Qualifiers  
 1. 416  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="srpb2-00055-cl1"  
 /tissue\_type="prostate tissue"  
 /clone\_1ib="srpb2 (10220)"  
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; rat  
 prostate normalized double selected poly(A+) mRNA size  
 fraction > 1 kb"

**ORIGIN**

**Query Match** 3.8%; Score 58.4; DB 6; Length 416;  
 Best Local Similarity 60.9%; Pred. No. 0.019;  
 Matches 95; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

532 AGAGCAGTGTGCGGTGTGTGCGAGCGCGGAGCTATCTGTGTGACGGCTGCTTC 591  
 14 AGGACTATATGAGGAGGTGTGCGAGCAAGCGGCGAGATCATCTGTGTATCTGTGCCCC 73  
 592 GGGCCCTTCACCTGAGCGCTGCTGCTCCCGCTCCGCGGAGATCCCAAGTGGAGCTGGA 651  
 74 GAGCTACCATATGTGTGCTGTGAGCCAGACATGAGAGAGCCCGGAGGCGAGTGA 133  
 652 GGTGTCTCAGCTGCTGCGAGCAAGCTCCAGAG 687  
 134 GCTGCCCACTGTGAGAGAGAGGATTCATCGGG 169

**RESULT 89**  
**LOCUS** CNS00720/c 932 bp DNA linear GSS 03-JUN-1999  
**DEFINITION** Drosophila melanogaster genome survey sequence T7 end of BAC #  
 BAC114B09 of RP11-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
**ACCESSION** CNS00720  
**VERSION** AL066742  
**KEYWORDS** AL066742.1 GI:4945205  
**SOURCE** GSS.  
**ORGANISM** Drosophila melanogaster (fruit fly)  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
**REFERENCE** 1 (bases 1 to 932)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - FRANCE (E-mail : segreifgenoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.flyfritzy.org/The\\_BDGP\\_Drosophila\\_melanogaster\\_BAC\\_library.html](http://www.flyfritzy.org/The_BDGP_Drosophila_melanogaster_BAC_library.html)

FEATURES

source

1..932  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BACR14B09"  
/clone\_lib="RPCL-98"  
/note="end : T7"

ORIGIN

Query Match 3.8%; Score 58.4; DB 9; Length 932;  
Best Local Similarity 31.0%; Pred.No. 0.021; Mismatches 162; Indels 0; Gaps 0;

MATCHES

109; Conservative 81; Mismatches 162; Indels 0; Gaps 0;

544 CCGTGTGCGGGAGCAGCGGAGAGCTCATCTGCTGAGACGGTGCCCTCGGCTTCACAC 603  
|||:::||||:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
861 SCGCCSSSSCGCGCCGCCSGGGGGCCSSSGSCSSGCGCCSSGCGCCSSSSSSCGCS 802  
|||:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
604 TGCGCTGCCTGTCCCTTCGCTCCGGAGATGCCAATGGAGACTGGAAGTGTCAACT 663  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
801 CSSCGCGSSSGCGSCCGCGSCGCGSCCGSGCGSGCGSGCGSCCGSCSSSGCGSGG 742  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
664 GCCTTCAGAGCAACATGCCAGAGGTGACAGCCCCGGGACAGAGAGCCCGCCCAAGAG 723  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
741 SGGGGSGSGSGGCGSSSGCGSSGCGCGGSCCGCGCGGGGGGCGSSGGGCGCGGS 682  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
724 CACCCTGAGAGACCCCCTGCCCTCCCGCGGGCTTAAGTGTGGCGGAGAGAGATTAGAAGTTC 783  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
681 SGCCGCGSCSCCSSGCGSCGCGCGCGCGCGCGCGSGSGCGSCCGCGCGCGCG 622  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
784 CACCTGGGGAAACCTTAGCCGCGCAATGACAACGACTTTGTTCMAACACACTGCGCGGCTC 843  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
621 SCGCCCCCGCGSCSGAGSGSCCGCCMCMAAMAVSVSCCCCGCMASGCCGCVSGGSC 562  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
844 CGCCTTGTGACAGCCCGCTGCGCAGAGTGTGACTCCTCGGCGCTGACCCCTT 895  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||  
561 CSCSMSCCCCCGCGCASCVCAGCGCMASMGCGCGCGCCGCGCASCSCGCTCCY 510  
:::||||:::||||:::||||:::||||:::||||:::||||:::||||

RESULT 90

LOCUS BP761124

DEFINITION BP761124 mouse (C57BL/6) pancreatic islet library with recombination-based method Mus musculus cDNA clone m1c01051 5', mRNA sequence.

ACCESSION BP761124

VERSION BP761124.1 GI:50219822

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 302)

Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H., Takeda,J., Ohara,O. and Seino,S.

TITLE	Construction of a multi-functional cDNA library specific for mouse									
JOURNAL	Published (2004)									
COMMENT	Contact: Suenmu Sello Division of Cellular and Molecular Medicine Kobe University Graduate School of Medicine 7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan Tel: 81-78-382-5360 Fax: 81-78-382-5370 Email: sellom@med.kobe-u.ac.jp.									
FEATURES	Location/Qualifiers									
SOURCE	1..302									
	/organism="Mus musculus"									
	/mol_type="mRNA"									
	/strain="C57BL/6"									
	/db_xref="taxon:10090"									
	/clone="mic01051"									
	/sex="male"									
	/tissue_type="pancreatic islet"									
	/dev_stage="adult"									
	/clone_lib="mouse (C57BL/6) pancreatic islet library with recombination-based method"									
ORIGIN										
Query Match	3.8%; Score 58.2; DB 5; Length 302;									
Best Local Similarity	58.3%; Pred.No. 0.021;									
Matches 102; Conservative	0; Mismatches 73; Indels 0; Gaps 0									
Oy	532	AGAGCAGATGTCCTGTCGTGTCGGACAGCGCGGGAGCTCATCTGCTGTACAGCGCTGCCCC	591							
Db	77	AGGATTCCTGTGAGAGTGTGCCAGCAGCGCGGGGAATATTCGTGTGCCACACTGCCCC	136							
Oy	592	GAGCGCTTCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGACCTGGA	651							
Db	137	GTCGCTACCACTGTGTGTGCTTGACCCCGAGCTTGACCGGCTCTGAGGCGCAATGGA	196							
Oy	652	GGTGTCTCAGTGCCTGCAAGCACAATTCACAGAGTGTGACGCCCCGGGCAGAGGA	706							
Db	197	GCTGCCCCCACTGTGAGAGAGGAGGGGTGAGTGGAGGCCAAGAGAGAGAGA	251							
RESULT 91										
CF535771	357 bp mRNA linear EST 12-SRP-200									
LOCUS	UI-M-GH0-cha-c-05-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone									
DEFINITION	IMAGE:30533716 5', mRNA sequence.									
ACCESSION	CF535771									
VERSION	CF535771.1 GI:34587739									
KEYWORDS	EST.									
SOURCE	Mus musculus (house mouse)									
ORGANISM	Mus musculus									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
REFERENCE	1 (bases 1 to 357)									
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .									
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)									
JOURNAL	Unpublished (1999)									
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at <a href="http://genome.uiowa.edu/distribution/mouse1.html">http://genome.uiowa.edu/distribution/mouse1.html</a> This clone was contributed by the Brain Molecular Anatomy Project (BMAP)									
FEATURES	The following repetitive elements were found in this cDNA sequence: 302-357, >(GGA)nSimple_repeat									
Seq primer: PYX-5.	Location/Qualifiers									
1..357	/organism="Mus musculus"									
SOURCE										

/mol\_type="mRNA"  
 /strain="CS7BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30533716"  
 /tissue\_type="whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (71 phage resistant)"  
 /clone\_1fb="NH\_BMAP\_G10"  
 /note="Organ: Brain; Vector: pYX-Asc; Site\_1: EcoR I; Site\_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGCTAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

```

organism="Mus musculus"
mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30533726"
/tissue_type="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phase resistant)"
/clone_1ib="NH BMAP.GHO"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CCAATCGATT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemmi Chin, Ph.D.,
Program coordinator."

```

Query Match	3.8%	Score 58.2;	DB 7;	Length 361;
Best Local Similarity	58.3%	Pred. No. 0.021;		
Matches 102; Conservative	0;	Mismatches 73;	Indels 0;	Gaps 0

[illegible]

LOCUS	428 bp	mRNA	linear	EST 27-FEB-1998
AA839768				
AA839768				
vw51d11.t1	Soares_mammary_gland_NBMNG	Mus musculus	CDNA clone	
DEFINITION				

KDA PROTEIN ; , mRNA sequence.

VERSION AA839768.1 GI:2916362  
KEYWORDS EST.

REFERENCE  
AUTHORS

Eukaryote, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.  
1 (bases 1 to 428)

Merra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

TITLE	JOURNAL	COMMENT
The Washu-HHMI Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse EST Project Washu-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Email: mouseest@watson.wustl.edu

1. .361

IMAGE Consortium (info@image.llnl.gov) for further information.



Seq primer: M13 Reverse  
High quality sequence stop: 575  
POLYA=No.

# FEATURES

source

Location/Qualifiers  
1. 575  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6Ncr"  
/db\_xref="taxon:10090"  
/db\_xref="taxon:10090"  
/clone="NIA:A0615D04 IMAGE:30749031"  
/dev\_stage="Age ~10 weeks old"  
/lab\_host="DH10B"  
/clone\_11b="NIA Mouse Hematopoietic Stem Cell  
(lin-/c-Rit-/Sca-1+) cDNA library (long 1)"  
/note="Vector: PCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://jgen.gsc.nia.nih.gov/cDNA).  
This is a long-transcript enriched cDNA library (Ref.  
Genome Res. 11: 1553-1558 (2001). [PMID: 11544193]). Total  
RNAs were obtained from Drs. Dennis Taub, Dan Longo  
(National Institute on Aging, USA), Jonathan Keller  
(National Cancer Institute, USA). Double-stranded cDNAs  
were synthesized with an Oligo(dT) primer (Invitrogen):  
5'-pGACTGTTCTAGATCGACGCGCGCTTTTCTTTT-3' from  
1.1 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to lone-linker L1-SalI, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer SalI-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of PCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 2.2 Kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 3.8%; Score 58.2; DB 7; Length 575;

Best Local Similarity 58.3%; Pred. No. 0.022; Mismatches 73; Indels 0; Gaps 0;

Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

532 AGGACGAGTGTGCGCTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGCTC 591  
173 AGGATTACTGTGAGGTGCGACAGGGGGGGAATTATTCGTGCGACACTGCCCC 232  
592 GGGCCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCAAGTGGAGCTGGA 651  
233 GTGCTTACCACTCTGCTGCTTGAACCCGAGCTTACCGGGGCTCTGAGGGGCAAGTGA 292  
652 GGTGCTCCAGCTGCTGCTGCGACGACAGTCCAGAGGTGACGCCCCGGGCGAGAGA 706  
293 GCTGCCCCCACTGTGAGAGAGAGGGGGTGCAGTGGAGGCCAAGAGAGAGAGA 347

RESULT 96  
LOCUS BUT03948 690 bp mRNA linear EST 15-JUL-2003  
DEFINITION UI-M-F00-btr-e-16-0-UI-r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
IMAGE:6406263 5', mRNA sequence.  
ACCESSION BUT03948  
VERSION BUT03948.1 GI:23631529  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (baase 1 to 690)  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

## JOURNAL

COMMENT

Unpublished (1999)  
Contact: Robert Straube, Ph.D.  
Email: c5abbs-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

The following repetitive elements were found in this cDNA  
sequence: 1-42, >(CAG)n\$Simple\_repeat 647-650,  
>(GGA)n\$Simple\_repeat  
Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers  
1. 690  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6406263"  
/issue\_type="whole brain"  
/dev\_stage="embryo 12.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_11b="NIH BMAP\_F00"  
/note="Organ: Brain; Vector: pYX-Anc; Site 1: BclI;  
Site 2: NotI; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a NotI site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with BclI adaptor, digested with NotI and then cloned  
directionally into pYX-Anc vector. The library tag  
sequence located between the NotI site and the polyA tail  
is TAGAGAGCC. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

## ORIGIN

Query Match 3.8%; Score 58.2; DB 5; Length 690;

Best Local Similarity 58.3%; Pred. No. 0.022; Mismatches 73; Indels 0; Gaps 0;

Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

532 AGGACGAGTGTGCGCTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGCTC 591  
513 AGGATTACTGTGAGGTGCGACAGGGGGGGAATTATTCGTGCGACACTGCCCC 572  
592 GGGCCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCAAGTGGAGCTGGA 651  
573 GTGCTTACCACTCTGCTGCTTGAACCCGAGCTTACCGGGGCTCTGAGGGCAATGGA 632  
652 GGTGCTCCAGCTGCTGCTGCGACGACAGTCCAGAGGTGACGCCCCGGGCGAGAGA 706  
633 GCTGCCCCCACTGTGAGAGAGAGGGGGTGCAGTGGAGGCCAAGAGAGAGAGA 687

RESULT 97  
LOCUS AK045449 2253 bp mRNA linear HTC 03-APR-2004  
DEFINITION AK045449 full-length enriched library, clone:B23020M16 Product:HT-2  
autantigen 240 kDa protein (fragment) homolog [Homo sapiens], full  
insert sequence.  
ACCESSION AK045449  
VERSION AK045449.1 GI:26337378  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus (house mouse)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meeth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsuno, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076661

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2253)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hangaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, I., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-Jun-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, FAX: 81-45-503-9216]

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.  
Location/Qualifiers  
1. .2253  
/organism="Mus musculus"  
/mol\_type="mRNA"

FEATURES  
source

	CDS	/getrahn="C57ML/SJ" /db_xref="FANTOM_DB:B23020LM16" /db_xref="taxon:10090" /clone="B23020LM16" /sex="male" /tissue_type="corpora quadrigemina" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" 129_->2252 /note="unnamed protein product; Mi-2 autoantigen 240 kDa protein (fragment) homo [Homo sapiens] (P1R1I8556, evidence: FASRY, 96.8%id, 98.4%length, match=1578) putative" /codon_start=1 /protein_id="BAC32375.1" /db_xref="GI:26337379" /translation:"MRRTKREYSAGCPGHSRDNRSPSCHLPFRPPPLDPRLD KDSDIRLPALCVKRKKRPQKEKPPEPKRKKLDSSEFGSERDYRESBSG VSEBGTPGRKRRKHNEKKETKTREBGSGNQNVSATLLTWLEDSVA VFSEEDYHTLTNYKAIFSOFWRPLIAKNPKIPMSKMWTIGAKMERPSANPRGSNA AVAAHAALAAAAAVALVEQSVAIVSATPIAPESGPPLAPPAPIORPIRAKTEKG GPSHDNRNKNSPRVPDGKKLRGKKMAPLTKILGLGGKKKASCAROSEGHDEPAE ESPIDSSSYSAAGMPGPGPYARKTARGRGKRVKKYLGAATGEENGVLENDDH YCYRCQQQGEITLCDCTPPRAYHLVCLDPELLDRAPBGWCSPHKEKVQCBAAZEEB YESEBEGEREBEEDDHMEYCVCCKDBELLCDAICISYTHICLNPLLPDI PNEGWLJC PRCTCPVLTRGVOKILHMWEBPFPVAIPAQAODMDVPPRLPQRSEREPFKVVNV GLSTWCSTAKEELQLIEFLHWMTARYORKNDMEPPLDYGSGSDDSKSRKVKQPH YAEMEEKYRFPGIKPEMWTATHRIINSHMDKKGNHYLVWKMDLPYPDSTWEBDEMNP EYVDHKQSVVRHRELIMGEDPAQRK"
Query Match	3.8%; Score 58.2;	DB 3 ; Length 2253; Best Local Similarity 58.3%; Pred. No. 0.026; Matches 102; Conservative % 0; Mismatches 73; Indels 0; Gaps 0;
Oy	532 AGGACGAGTGTCGCCTGTGCCGGAGCGCGGGGAGCACTGCTGTGACGAGCCCTC	591
Db	1300 AGGATTACTGTGAGAGTGTGCCAGAGGACGGGGAATAATTATTCTGTGGACA	1359
Oy	592 GGAGCTTCCACTGGCTGTGCCCTGTCCCCTCCGCTCCGGAGAATCCCGAGTGGAC	651
Db	1360 GTGCTTAACCACCTGTCTGCTTGAACCCCGAGCTTAGACGGGGCTCTGAGGGCAAGTTGA	1419
Oy	652 GGTCTCTCACGTGCTCTCGACGCAACACTCCAGAGAGTGTGACGCCCCGGGACAGGA	706
Db	1420 GCTCCCTCCCATCTGTGAAGAAGGAGGGGGTGTGAGTGGAGGCCAACGAGAGAGGA	1474
RESULT 98		
CNS0072Q		
LOCUS		
DEFINITION	CNS0072Q 932 bp DNA linear GSS 03-JUN-1999	
FEATURES	drosophila melanogaster genome survey sequence T7 end of BAC #	
BACR1AB09 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
AL066742		
AIO66742.1 GI:4945205		
GSS.		
Drosophila melanogaster (fruit fly)		
Drosophila melanogaster		
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
1 (bases 1 to 932)		
Genoscope.		
Direct Submission		
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
BP 191 J1006 EVRY cedex - FRANCE (E-mail : seqrete@genoscope.cns.fr		
- Web : www.genoscope.cns.fr)		
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila		
COMMENT		

melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; on bw sp. the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

```
1..932
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14B09"
/clone_lib="RPCL-98"
/notes="end : T7"
```

## ORIGIN

```
Query Match      3.8%; Score 58; DB 9; Length 932;
Best Local Similarity 31.3%; Pred. No. 0.026;
Matches 102; Conservative 77; Mismatches 147; Indels 0; Gaps 0;

QY 859 CGCTGCAGGTCGACCTCTCGGCTGACACCCCTACTGTGTGAGGCTCTGAGGCTC 918
DB 531 CGGKGTGKGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
QY 919 AGCAGAACCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
DB 591 YTBKAGGCGGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 650
QY 979 GGTGTACTACTGCGCGCTGCTCTTCACTGCGCTGCTGCTGCTGCTGCTGCTGCT 1038
DB 651 GCGCCCGGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 710
QY 1039 GCGCGGAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1098
DB 711 GCGCGGAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 770
QY 1099 AGGAGGCTGCTGCGCCCGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1158
DB 771 CGGCGGAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 830
QY 1159 CCAGTACAGACCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
DB 831 CCGGAGAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
```

## RESULT 99

AG057392 997 bp DNA linear GSS 02-NOV-2001  
 DEFINITION Pan troglodytes DNA, clone: PTB-043P11.R, genomic survey sequence.

AG057392  
 AG057392  
 AG057392.1 GI:16594851  
 GSS.  
 SOURCE Pan troglodytes (chimpanzee)  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
 Totoki, Y., Watanabe, H., and Sakaki, Y.

TITLE BAC end sequences of Library PTB  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 997)  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suenho-chou, Tsukuba, Ibaraki, Japan  
 (E-mail: [chimpes@gsc.riken.go.jp](mailto:chimpes@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>),

## COMMENT

Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB. This BAC end  
 was generated during the Rad process and may have higher chance of  
 clone tracking errors.

## PRIMERS

Sequencing: M13rev  
 LIBRARY  
 Vector : pKS145  
 R.site 1 : SacI  
 R.site 2 : SacI.

## FEATURES

source

```
1..997
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-043P11.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
```

## ORIGIN

```
Query Match      3.8%; Score 58; DB 9; Length 997;
Best Local Similarity 44.6%; Pred. No. 0.026;
Matches 229; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 696 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 755
DB 121 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 756 AGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 815
DB 181 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 816 ACTTGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
DB 241 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 876 TCTTGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 935
DB 301 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 936 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 995
DB 361 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 996 GCTGCTTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055
DB 421 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 1056 CGTGTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1115
DB 481 CCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 1116 AGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1175
DB 541 CCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 1176 CTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
DB 601 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 634
```

## RESULT 100

AG057392 1038 bp mRNA linear EST 21-AUG-2002  
 DEFINITION IMAGE:6198951 5', mRNA sequence.

AG057392  
 AG057392  
 AG057392.1 GI:22369403  
 GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE  
1 (bases 1 to 1038)  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.jnl.gov>  
Plate: LIML3611 row: 9 column: 16  
High quality sequence stop: 214.  
Location/Qualifiers

## FEATURES

source

```

1..1038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6198951"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH108"
/clone_lib="Lupski_sciatic nerve"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACTAGCTTCAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

```

## ORIGIN

```

Query Match      3.7%; Score 57.8; DB 5; Length 1038;
Best Local Similarity 37.9%; Pred. No. 0.029;
Matches 307; Conservative 0; Mismatches 498; Indels 6; Gaps 2;

QY 290 CCGGGTTTCTTCCCATAGGATGCCCCGGGGGTCTCTTCGAGACCAATGATGATG 349
DB 195 CCAAGCTTGAAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 254
QY 350 GGGACAGGTGATCAGGGCAAAATTTCAAGCCCTGACATGGAGAGAGGACGACT 409
DB 255 GNGNNGGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 314
QY 410 GGGAGTTCAAGTACCAGAG-ATGCTGCTGGGGGAGCTTTTGGAAAGAGTGGCTC 468
DB 315 GGGGGGGCCGGGNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 374
QY 469 TCAGAGGGGTGTCACCCAGCCAGTCTGATGGGCGTCTTGTGCTGTGCAAGA 528
DB 375 CANNNGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 434
QY 529 ATGAGAGCGAGTGTGCTGTGTGAGGACGCGGGAGCTCATCTGCTGTGACGCTGCC 588
DB 435 CCCCCCGGGCGGNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 494
QY 589 CTGGGGCTTTCCACTGTGGCTGTGCTGCTGCTGCGGGAGATCCCAATGGAGCT 648
DB 495 CCNGGGCCCGCCCCCGCCCCCGCCGAGGGGAGNGGCGGNGCGCCCCCGGGGG 554
QY 649 GGAGGTCTTCAGTCTGCTGAGGCAAGTCCAGGAGGTGACCCCGGGGAGAGAGC 708
DB 555 GCGGGGGCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 614
QY 709 CCGGGCCCCAGAGGACCCCGTGAAGACCCCGCTCCCGGGGGCTTAAGTCGGCGGAG 768

```

```

DB 615 CCGGGGGGGGGGGCGCCCGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 674
QY 769 AGAGGTAAAGAGTCCACTTGGGGAACCCCTTAGCGGCAATGACAGACTTTGTATA 828
DB 675 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 734
QY 829 AGACCTGGCG-----GTCGGCTTTCTGACAGCCCGCTGCCAGGTCTGAATCTCTCGAC 883
DB 735 NCCNNNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 794
QY 884 CCTGACCCCTTAATGTGTGTGTGAGGTCTGAGGCTCAGAGAACTGGACTCTGTGGCGG 943
DB 795 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 854
QY 944 TTGGGGGTGTGCGGAGATGTATCGGAGCTGTGGGTATCTACTGTGCGCGCTGCTT 1003
DB 855 NCGCGCGGNNCCCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 914
QY 1004 CCACTGGCGCTGCACCTTCCAGCGGACCTCCCGGCGGGGAGCGGGCTGCTGACG 1063
DB 915 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 974
QY 1064 ATCTGCTCAGAGAGCGTGAACCCCAAGCCCT 1094
DB 975 CTCGCGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1005

```

Search completed: February 9, 2005, 17:40:05  
Job time : 4565 secs